Published in final edited form as:
*Proteins*. 2011 November ; 79(11): 3166–3179. doi:10.1002/prot.23149.

# Design of a Modified Mouse Protein with Ligand Binding Properties of its Human Analog by Molecular Dynamics Simulations: The Case of C3 Inhibition by Compstatin

**Phanourios Tamamis**[a,c,b,$], **Panayiota Pierou**[a], **Chrystalla Mytidou**[a], **Christodoulos A. Floudas**[c,*], **Dimitrios Morikis**[b,*], and **Georgios Archontis**[a,*]

[a]Department of Physics, University of Cyprus, PO20537, CY1678, Nicosia, Cyprus.

[b]Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, USA.

[c]Department of Bioengineering, University of California, Riverside, California 92521, USA.

## Abstract

The peptide compstatin and its derivatives inhibit the complement-component protein C3 in primate mammals and are potential therapeutic agents against unregulated activation of complement in humans, but are inactive against C3 from lower mammals. Recent molecular dynamics (MD) simulations showed that the most potent compstatin analog comprised entirely of natural amino acids (W4A9) had a smaller affinity for rat C3, due to reproducible changes in the rat protein structure with respect to the human protein, which eliminated or weakened specific protein-ligand interactions seen in the human C3:W4A9 complex. Here, we study by MD simulations three W4A9 complexes with the mouse C3 protein, and two “transgenic” mouse derivatives, containing a small number (6-9) of human C3 substitutions. The mouse complex experiences the conformational changes and affinity reduction of the rat complex. In the “transgenic” complexes the conformation remains closer to that of the human complex, the protein-ligand interactions are improved, and the affinity for compstatin becomes “human-like”. The present work creates new avenues for a compstatin-sensitive animal model. A similar strategy, involving the comparison of a series of complexes by MD simulations, could be used to design “transgenic” sequences in other systems.

### Keywords

Innate immune response; Complement system; C3 inhibitors; Compstatin analogs; Molecular Dynamics

## 1. Introduction

The complement system provides the first line of defense against the invasion of foreign pathogens [1]. The role of the component system involves opsonization, inflammation and lysis and includes autoimmunity, clearance of immune complexes, debris removal and response to tissue injury. Inappropriate activation of complement may cause or aggravate several pathological conditions, such as asthma, adult respiratory distress syndrome, hemolytic anemia, macular degeneration, rheumatoid arthritis, rejection of

*Corresponding authors: Georgios Archontis: archonti@ucy.ac.cy; Dimitrios Morikis: dmorikis@engr.ucr.edu; Christodoulos Floudas: floudas@titan.princeton.edu.
$Phanourios Tamamis is a Fulbright visiting Scholar at the University of California, Riverside and at Princeton University.

xenotransplantation, stroke and heart attack [2,3,4,5]. Therefore, the development of drugs for the control of complement activation is of profound interest.

Complement action proceeds via three biochemical pathways (classical, alternative and lectin), which converge to a common point, the cleavage of protein C3 to fragments C3b and C3a. The large fragment, C3b, tags pathogen surfaces for recognition by phagocytic cells (opsonization), and the small fragment, C3a, aids in immune cell recruitment (chemotaxis) and inflammation. The C3b fragment also participates in complexes, called convertases, which are responsible for cleavage of C3 to C3a and C3b, as well as cleavage of complement protein C5 to C5a and C5b. C5 is the starting protein of the common activation pathway, which ends with the formation of the membrane attack complex (MAC), a multicomponent complex involved in lysis of pathogen membranes. Protein C3 is essential in all pathways and represents a good target for complement inhibition. For example, regulation of C3 cleavage would control the effects of C3a and C3b, and the progression of complement activation to C5, and, eventually, to MAC. Altogether, regulation of C3 would affect the opsonization, chemotactic, inflammatory, and lytic capabilities of the complement system.

The 13-residue compstatin inhibits the cleavage of human C3 (hC3) to C3a and C3b by binding to the C3 β-chain [4, 6, 7] and is a promising candidate for the therapeutic treatment of unregulated complement activation. Compstatin also binds to the C3b fragment as well as the inactive C3c fragment, both of which contain the β-chain. Compstatin was first discovered by a phage-displayed random peptide library for binding against C3b [8]. Subsequent structural (NMR) experiments determined the thermodynamically dominant conformation of various compstatin analogs in solution, with residues 5-8 in a type I β-turn and terminal residues 1-4/12-13 in a hydrophobic cluster [9, 10, 11, 12, 13]; molecular dynamics simulations investigated further the solution properties of compstatin analogs [14, 15], mutational studies identified aminoacids critical for C3 inhibition [12, 13, 16, 17, 18, 19] and computational studies attempted to relate the inhibitory potential with specific physicochemical properties of the peptide [20, 21]. Furthermore, rational design, experimental combinatorial design and computational combinatorial design methods suggested a number of compstatin analogs with natural or artificial aminoacids [6, 12, 13, 16, 18, 19, 22, 23, 24, 25, 26, 27, 28, 29].

A breakthrough in the understanding of the conformational properties and interactions of the human C3:compstatin complex was achieved by a recent X-ray crystallography study, which determined the structure of the complex between the proteolytic fragment C3c of C3 and the compstatin analog W4A9 at 2.4 Å resolution [4]. C3c is convenient for co-crystallization studies as it maintains the structural characteristics of C3 and C3b at the β-chain level, which contains the compstatin binding site, and offers the advantage of a smaller size fragment than C3 and C3b. In addition, C3c is free of the C3d domain (site of the opsonization thioester bond), which is the activation domain of intact C3. Compound W4A9 is the N-terminal acetylated (Ac) double mutant Ac-Val4Trp/His9Ala (W4A9). It is more active than native compstatin by 45-fold (15-fold, compared to a native variant, acetylated at the N-terminal end) [11]. W4A9 binds at the protein surface and forms extensive interactions with four segments in the MG4 and MG5 domains [4] by adopting a conformation that is significantly different from the thermodynamically dominant conformer of native compstatin in solution. The β-turn in segment 5-8 is replaced by a new β-turn in the region 8-11; the average backbone $C_\alpha$-atom root-mean-square difference (RMSD) between the conformation of W4A9 in the complex and the thermodynamically dominant conformation of native compstatin in solution is 3.7 Å.

The development of effective compstatin inhibitors against non-primate C3 can be used to test disease models in non-primate animals. An important conclusion, drawn from experimental studies with a large number of species, was that compstatin is active against C3 from several primate species but fails to inhibit the activation of C3 from lower mammals [8, 30]. The experiments suggested that the lack of activity against non-primate C3 was not due to protein-ligand steric or electrostatic repulsion, but due to the loss of specific interactions and/or structural rearrangements of the proteins [30]. This conclusion was supported by a recent computational study, in which we simulated complexes between W4A9 and human or rat (*rattus norvegicus*) C3 [31].

Rat C3 (rC3) is highly homologous to its human counterpart; the compstatin binding region (residues 329-534 of segments MG4-MG5 [4]) has ~80 % homology in the two proteins. Thus, the rat and human C3 conformations are expected to be very similar. This was indeed observed in the simulations [31]; at the same time, certain protein segments in proximity to W4A9 underwent small-scale conformational changes, causing the weakening or loss of key hydrogen-bonding interactions and non-polar contacts with the ligand. This behavior was reproducible in simulations of the mouse C3:W4A9 complex (this work). In contrast, the simulation conformation and interactions of the human C3:W4A9 complex [31] were similar to those of the crystallographic complex [4].

We can use the insights obtained from the MD simulations [31] to suggest potentially active inhibitors against non-primate mammalian C3, or improved inhibitors against human C3. We can pursue the first objective in two ways. We can modify the ligand by applying *de novo* design methods [27, 29] on structural models from the MD simulations of non-primate complexes. Alternatively, we can redesign the non-primate proteins with the aim to stabilize the conformation and/or protein-ligand interactions of the human complex [4, 31]. A simple and plausible choice would be to create a “transgenic” protein by inserting human C3 substitutions at key locations of the non-primate protein, which increase the similarity between the two binding sites.

This second path is promising for several reasons. First, it validates the use of experimental information from the human complex [4] in the “transgenic” design. Indeed, since the ligand is invariant and the “transgenic” and human proteins are highly homologous, we could expect that a successful “transgenic” complex would resemble structurally the human complex; thus, we could search for modifications that would recover the protein-ligand interactions seen in the human complex. Furthermore, if the resulting “transgenic” protein acquires a measurable affinity for W4A9, it is likely that it would also bind other compstatin-based human-C3 inhibitors [6, 12, 13, 16, 18, 19, 22, 23, 24, 25, 26, 27, 28, 29]. Conversely, novel inhibitors of the “transgenic” protein would be presumably active against human C3 as well, since the two proteins would share very similar binding sites.

In what follows, we detail our efforts along this latter direction. In a separate, forthcoming publication we will present simulations of various non-primate and human C3 complexes with a series of novel compstatin analogs. Here we focus on mouse (*mus musculus*) C3 (mC3), instead of rat C3 [31], since mouse genetics is much better studied. Simulations of the mC3:W4A9 complex show localized structural changes near the ligand, which cause the weakening or loss of important intermolecular interactions, as in the rat complex [31]. This reproducible behavior corroborates our interpration of the compstatin species specificity in terms of structural changes in non-primate C3 [31], and is a first result of the present study.

A main objective is to identify a small number of substitutions, which assist mC3 to retain a “human-like” structure and affinity for compstatin. This is not a priori obvious from the human/mouse C3 sequence alignment (Fig. 1). For example, an important difference

involves primate C3 residues His392/Pro393, which become Pro392/Asn393 in non-primate C3. In Ref. [31], these mutations were thought mainly responsible for the displacement of rat C3 sector 388-393, and the concomitant reduction for W4A9 affinity. As explained in the present work (*Discussion* section), test simulations show that the two substitutions 392-393 were not sufficient to retain the segment 388-396 near its crystallographic position. Our simulations suggest that the introduction of 6-9 human C3 substitutions in three mouse C3 sectors can restrict the conformational changes of the mouse/rat C3 complex and improve the protein-ligand interactions relative to the original mouse complex. Interestingly, several interactions are also improved relative to the human complex, and the protein acquires a human-like compstatin affinity. Thus, the present MD-based design suggests that a "transgenic" mouse-human protein with a human-like compstatin affinity is plausible. This type of work is in general useful for aiding in the design of transgenic animal models for experimental testing of inhibitors possessing species specific binding and inhibition properties.

## 2. Methods

### Choice of "transgenic" sequences

We simulated complexes of W4A9 with the mouse C3 (mC3) sequence, or two "transgenic" (mouse with human insertions; mhC3) variants of mC3. The first variant (mhaC3) contained six substitutions (Pro392His, Asn393Pro, Arg395Gln, Gln396Lys, His454Leu, Pro459Arg), and the second variant (mhbC3) three more (Ala345Gly, Thr457Met and Gly460Ala). All substitutions are indicated with bold letters in Fig. 1, which shows an alignment of the compstatin-binding region (domains MG4-MG5 with residues 329-534) of the human (hC3) and mouse (mC3) sequences, along with the two "transgenic" sequences mhaC3 and mhbC3. In what follows, we explain the reasoning behind our choice of substitutions.

The rectangular boxes in Fig. 1 indicate the four sectors (I-IV), which contact directly the ligand in the human complex hC3:W4A9 [4]. Sector I (residues 344-349 in hC3) is conserved across species, with the exception of position 345, which contains a glycine residue in primate and an alanine in non-primate mammals [31]. In the crystal structure and the simulations of the hC3:W4A9 complex, the carbonyl group of Gly345 formed a stable hydrogen bond with the main-chain amide group of the ligand Trp4. In the simulations of the rat C3 complex with W4A9, the occupancy of this hydrogen bond is somewhat reduced [31]; nevertheless, this mutation is probably not critical for affinity. For this reason, the substitution **Ala345Gly** is only tested in the second "transgenic" sequence (mhbC3).

Sector II (residues 388-393) forms hydrogen-bonding interactions with the N-terminal residues Ile1 and Cys2 of W4A9 [4,31]. In the simulations of rC3:W4A9 complexes [31], as well as the mouse simulations reported below, sector II is consistently displaced away from the ligand, with a concomitant loss of protein-ligand interactions. Thus, this sector was deemed important for the ligand affinity [31]. In the mC3 and rC3 proteins, His392 becomes proline and Pro393 becomes asparagine. Other residues of sector II are preserved among primate and non-primate sequences, with the exception of Ile389 (valine in rC3) and Ser388 (threonine in *Canis Familiaris*). To render sector 388-393 human-like, the substitutions **Pro392His** and **Asn393Pro** are included in both "transgenic" sequences, together with two additional substitutions at the C-terminal side of the sector (**Arg395Gln** and **Gln396Lys**). With these four changes, the segment 382-406 acquires an identical composition in the human and "transgenic" sequences (Fig. 1), with the exception of position 399, which is not invariant but retains its chemical similarity across species (serine in human and threonine in "transgenic"/mouse C3).

Sector III (residues 454-462) has six non-conserved residues (454, 457 and 459-462) across species [31]. Among these residues, Leu454 and Arg459 form the most important interactions with the ligand in the human complex [31]. Leu454 participates in a hydrophobic cluster with Met346, Pro347 and Val3 of the ligand. In all non-primate species, Leu454 becomes histidine, disrupting non-polar contacts in the cluster and lowering ligand affinity. This was clearly seen in the rC3:W4A9 simulations of Ref. [31], where a contact between the ligand Val3, His10 sidechains and the His454 was lost and Val3 became more flexible. It is also observed in the mouse simulations (below). Arg459 becomes alanine in rat C3 and proline in most other non-primate mammal sequences [31], including mouse (Fig. 1). Even though Arg459 does not make direct electrostatic interactions with the ligand, its non-polar moiety makes frequent contacts with the sidechain of Trp7 in the human C3 simulations, in line with the X-ray structure [4]. This contact is reduced in the simulations of the rat complex [31] and the mouse complex (below). The two substitutions **His454Leu** and **Pro459Arg** are inserted in both "transgenic" sequences studied here. In sequence mhbC3 we include the two additional substitutions **Thr457Met** and **Gly460Ala**, which render the entire sector identical to the one in human C3. Finally sector IV (488-492) is conserved in most species and is identical in human, rat and mouse C3.

## Ligand

We simulated the double mutant Ac-Val4Trp/His9Ala (W4A9) with sequence $COCH_3$-Ile1-Cys2-Val3-Trp4-Gln5-Asp6-Trp7-Gly8-Ala9-His10-Arg11-Cys12-Thr13-$NH_2$. The disulfide bond Cys2-Cys12 was maintained by a disulfide patch of the CHARMM topology file. Titratable residues were assigned their most common ionization state at physiological pH (charged Asp6 and Arg11, neutral His10).

## Protein

We employed the truncated C3c system of Ref. [31]. Briefly, this system contained the entire two domains MG4 and MG5 (residues 329 - 534) of the compstatin binding site and the proximal segment 607-620. The resulting protein:ligand complex had approximate dimensions 40×53×56 Å. Compstatin bound on one side of the truncated model facing the solvent, and was at least 21 Å away from any protein atom omitted in the simulation [31]. Poisson-Boltzmann calculations [32] established that the electrostatic interactions between compstatin and any omitted C3 charged residues were sufficiently screened by solvent [31].

The complexes were immersed in a box of water molecules in the shape of a 89-Å truncated octahedron; overlapping water molecules were omitted. The total charge of the simulation systems was set to zero, by addition of appropriate numbers of chloride anions. The resulting system corresponded to a solution with 0 M salt concentration. Modeling of a physiological ionic strength (0.15M salt concentration), even though it is strictly more accurate, is not likely to modify our results. Indeed, the contribution to affinity due to specific electrostatic interactions was small and fairly uniform across complexes (see *Results*).

The resulting mouse system had 35780 atoms (3657 protein and ligand atoms); the two "transgenic" mhaC3/mhbC3 systems had 35779/35743 atoms (3670/3673 protein and ligand atoms).

## Force Field Specifications

The peptide atomic charges, van der Waals and stereochemical parameters were taken from the CHARMM22 all-atom force field [33], including a CMAP backbone φ/ψ energy correction [34] and indole parameters from Ref. [35]. Force field specifications and

simulation protocols were the same as in Ref. [31]. All simulations were conducted with the molecular mechanics program CHARMM, version c35a2 [36].

### Initial Coordinates

With the exception of loop 369-378, which contains a deletion at position 372 (Fig. 1), the initial positions of all other backbone heavy atoms in all (mouse and "transgenic") complexes were taken from the crystallographic structure of the human complex (PDB code: 2QKI) [4]. With this choice, we avoided introducing any *a priori* structural differences between the human and mouse/"transgenic" complexes. Thus, our simulations investigated to what extent the mouse and "transgenic" complexes were able to retain conformations near the one of the human C3:W4A9 complex.

The initial conformation of the loop 369-378 was constructed with the program Modeller [37] and had a root-mean-square difference (RMSD) of 1.39 Å from the corresponding conformation in the human C3 protein. The heavy atoms of invariant sidechains were initially placed at the corresponding coordinates of the human complex. The initial positions of mutated sidechains were modeled with the SCWRL4 program [38]. Hydrogens were positioned by the HBUILD algorithm of the CHARMM program.

The main-chain heavy atoms of an external protein shell, with atoms at least 20 Å away from any atom of compstatin, were harmonically restrained to their initial crystallographic positions in order to avoid structural deformations at the protein boundary due to the truncation. Segments 373-377 of the reconstructed loop (373-376 in transgenic and mouse proteins) were also harmonically restrained in all simulations. The structure was initially optimized by 150 energy minimization steps with the steepest-descent and adopted-basis Newton-Raphson (ABNR) algorithms.

### Simulations of Mouse and "Transgenic" complexes

In the case of the mouse complex mC3:W4A9 we conducted three independent runs (M1-M3). We also conducted two runs (MHA1-MHA2) with the first "transgenic" complex mhaC3:W4A9, and four runs (MHB1-MHB4) with the second "transgenic" complex mhbC3:W4A9.

The starting conformation of runs M1, MHA1 and MHB1 corresponded to the second molecule in the asymmetric unit of the human C3:W4A9 crystallographic structure (PDB entry 2QKI). After an initial minimization, the complexes were subjected to 280 ps of equilibration, during which any imposed harmonic restraints within 20 Å of the ligand were gradually removed, as described in Ref. [31]. Runs M2, MHA2 and MHB2 were started after extending for 50 ps the last equilibration phase of runs M1, MHA1, and MHB1, respectively. Runs M3, MHB3 were started after extending for 50 ps the equilibration phase of runs M2, MHB2, respectively. Run MHB4 was started after extending for 50 ps the equilibration phase of run MHB3. The mouse (M1-M3) and "transgenic" MHB1-MHB4 runs had a "production-phase" length of 8.75 ns; the runs of the first "transgenic" complex (MHA1-MHA2) had a production-phase length of 7 ns. In all cases, we employed the last 7 ns for the analysis.

### Simulations of free proteins and free ligand

We conducted additional simulations of the free mouse and transgenic proteins MHA, MHB. The runs employed the same force field and computational protocol of the complexes and had a duration of 7 ns. The initial protein conformations were the same as in the corresponding complexes. For the free W4A9 ligand, we conducted ten independent 2-ns runs, starting from conformations 11-21 of the NMR family 1A1P [9]. The Cys2-Cys12

disulfide torsional angle ($2C_\beta 2S_\gamma$-$12S_\gamma$-$12S_\beta$) was set to +90°; the same force field was employed as in the complex and free protein simulations, with the computational protocol of Ref. [15].

## Analysis of side-chain contacts

Probability-density maps of intermolecular side-chain contacts were computed with the WORDOM package [39]. Two side-chains were considered in contact if the distance of their geometric centers was smaller than 6.5 Å.

## Computation of association free energies

The association free energies were computed by the relation

$$\Delta G = G_{PL} - G_P - G_L \tag{1}$$

where $G_{PL}$, $G_P$ and $G_L$ denote, respectively, the free energies of the complex (PL), the protein (P), and the ligand (L). These free energies were computed in the Molecular Mechanics-Generalized Born/Surface Area (MM-GBSA) approximation [40]. In this approximation, representative coordinates of each state X (X=PL, P, or L) are extracted from the corresponding simulation trajectories. The corresponding free energies are computed from the relation

$$G_X = E_X^{\text{bounded}} + \underbrace{E_X^{Coul} + E_X^{GB}}_{=G_X^{\text{polar}}} + \underbrace{E_X^{vW} + \sigma S_X}_{=G_X^{\text{non polar}}} \tag{2}$$

The first term on the right-hand side of Eq. (2) describes the dependence of the internal energy on the molecular geometry (bond lengths, bond angles, torsional angles); the second term describes Coulombic interaction energies between the atomic charges of the molecule; the third term represents the electrostatic contribution to the solvation free energy, and is modeled by the GBSW generalized-Born approximation [41,42]. The next term describes van der Waals interactions; the final term describes non-polar contributions to the solvation free energy, assumed proportional to the solvent-accessible surface area, $S_x$, of the molecule. The proportionality coefficient (σ) of this term was set to 0.005 Kcal/mol/Å$^2$, as in the GBSW parameterization. Note that for the complexes studied here, the contribution from this last term to the obtained affinities was approximately constant among complexes; thus, the value used for σ affects the individual but not the *relative* association free energies.

Contributions due to the changes in the protein and ligand entropies upon association were omitted in the association free-energy calculations. These quantities can have large errors (Ref. [43] and references therein, for a detailed discussion) and partially cancel in the comparison of the three complexes. Furthermore, the objective of these free-energy calculations is to estimate whether the inserted mutations confer to the “transgenic” proteins a human C3-like binding affinity for compstatin. The MM-GBSA estimates for the W4A9 association free-energies of the human and rat proteins were −55 kcal/mol and −46 kcal/mol, respectively, in the approximation of identical complex and unbound structures [31]. The MM-GBSA estimate for the binding free-energy of the mouse C3:W4A9 complex is expected to be near the estimate for the rat C3:W4A9 complex. If the estimates for the association free-energies of the “transgenic” complexes are near the human value, the resulting (“transgenic” – mouse) relative free-energies will be ~ −9 kcal/mol. This is a reasonably large free-energy difference, which could suggest that the inserted mutations have conferred human-like affinity for W4A9 in mouse C3, regardless of the approximations

in the MM-GBSA formalism. As explained in the *Results* section, the mouse and "transgenic" estimates indeed support this conclusion.

Interaction energies between two groups of atoms (R and R') were computed by the relation

$$\Delta G_{RR'}^{\text{inte}} = \underbrace{\sum_{i\in R}\sum_{j\in R'}\left(E_{ij}^{Coul}+E_{ij}^{GB}\right)}_{=\Delta G_{RR'}^{\text{polar}}} + \underbrace{\left[\sum_{i\in R}\sum_{j\in R'}E_{ij}^{vW}+\sigma\sum_{i\in R,R'}\Delta S_i\right]}_{=\Delta G_{RR'}^{\text{nonpolar}}} \quad (3)$$

The first and second group of terms on the right-hand side of Eq. (3) describe, respectively, polar and nonpolar interactions between R and R'; In our calculations, R corresponded to a compstatin residue and R' to the entire protein model; alternatively, R was a protein residue and R' was the entire ligand. To compute the GB term in Eq. (3), we included all protein and ligand atoms and set the charges of atoms outside the two groups R and R' to zero. The last term contains the difference in the SASA value of groups R, R' in the complex and unbound states.

The generalized-Born energies and the atomic accessible-surface areas ($\Delta S_i$) entering in Eq. (3) depend on the location of R and R' in the complex. The polar component contains a Coulombic term and a GB contribution, modeling the interaction between group R and the solvent polarization potential induced by R'. Similarly, the nonpolar component contains a van der Waals interaction between R, R' and a surface term, expressing cavity contributions and nonpolar interactions with the surrounding solvent. The sum of the two components reflects the total direct interaction between R and R' in the solvated complex.

## 3. Results

### The Human C3:W4A9 Complex

The simulations of the hC3:W4A9 complex were conducted and analyzed in detail in Ref. [31]. Here we provide an overview of the results, to assist the reader in comparing the human, mouse and "transgenic" results. A representative conformation of the hC3:W4A9 complex (last snapshot of simulation H2 [31]) is shown in Figs. 2A-B. Figs. 3A-B plot intermolecular interaction energies for W4A9 and C3c residues, computed with Eq. (3).

In the crystal structure (PDB code: 2QKI) [4] and the MD simulations [31], the bound conformation of W4A9 is extended in the region 5-8 and contains a β-turn in the region 8-11 and a 3-11 β-bridge (see Fig. 2A and Supporting Information Fig. SF1). The ligand N- and C-terminal moieties make extensive hydrogen-bonding and non-polar interactions with residues in four hC3 sectors shown in Fig. 2A; 345-349 (sector I), 388-393 (sector II), 455 - 458 (sector III) and 488 – 491 (sector IV). Ligand residues Trp7, Trp4, His10, Gln5, Val3, Ile1, Cys2, Ala9 (Fig. 3A) and protein residues Asp491, Arg456, Asn390 and Gly345 (Fig. 3B) form strong intermolecular interactions (in decreasing order). Important hydrogen-bonds and non-polar contacts are illustrated, respectively, in Figs. 2A and 2B. Residues Ile1 and Cys2 form hydrogen bonds with the side chain of Asn390; the Val3 side-chain forms a stable hydrophobic cluster with Met346, Pro347, and Leu454; the Trp4 main-chain makes a very stable hydrogen bond with Gly345 CO and a somewhat less stable interaction with the sidechain of Arg456; its side chain packs against the Cys2-Cys12 disulfide-bridge on one side and makes a weak hydrogen-bond with Asn391 and nonpolar contacts with Pro393 and the Ca atom of Gly345 on the other side; the Gln5 side-chain makes two intermolecular hydrogen bonds with main-chain groups of Leu455 and Met457; the Trp7 side-chain intercalates between sectors 455–458 and 488–491, making a stable hydrogen bond with

Met457 CO and non-polar contacts with Gln5, Met457, Arg459, and Glu462; the main-chain NH groups of Ala9 and His10 form very stable hydrogen bonds with the Asp491 side chain. Finally, the His10 sidechain interacts with Asp491 and is near a hydrophobic nucleus formed by Leu454 and Leu492 [31].

## The Mouse C3:W4A9 Complex

We conducted three simulations (M1-M3), with a "production" length of 8.75 ns. As explained in *Methods*, to avoid building into the mouse model any *a priori* structural differences from the human complex we initiated the mouse simulations from the experimental conformation of the human complex. With this choice, our simulations investigated whether the mouse complex was able to maintain the conformation of the human complex, or had the propensity to undergo conformational changes with a concomitant decrease in ligand affinity, as in the rat complex [31].

**Simulation conformations—**Root-mean-square differences (RMSD) between the simulation conformations and the crystallographic conformation of the human complex [4] are reported in Table I. The total RMSD of the protein backbone (second column) is in the range 0.85-0.86 Å, near the corresponding values for the human complex. Thus, the mC3 conformation remains near the crystallographic structure of the human complex. Sectors I (344-349) and II (388-393) deviate more from the crystallographic conformation. This is illustrated in Fig. 2C, which displays the conformation of the compstatin binding site at the end of run M3, along with the initial conformations of these two sectors. A trajectory corresponding to run M3 is shown in Supporting Information Video M3.

The displacement of the mouse sector I is partly related to the presence of an alanine instead of glycine at position 345, which reduces the flexibility of the backbone and alters somewhat the conformation near residue 345. In the crystal structure, the Gly345 phi/psi torsional angles have values 98° and 16°, respectively [4]. In the MD simulations of the human complex [31] and the "transgenic" complex mhb, the corresponding Gly345 phi/psi average values are 86°/8° and 84° /11°, respectively. On the other hand, the Ala345 phi and psi torsional angles have average values 60°/38° in the mouse runs and 69°/35° in the first "transgenic" sequence mha. Thus, the experimental values are much closer the human (H) and second "transgenic" attempt (MHB) runs, compared to mouse (M) and the first "transgenic" attempt (MHA) runs.

Sector II is part of segment 388-396, which contains four mutations with respect to the hC3, in positions 392-393 and 395-396 (Fig. 1). Introduction of human residues at all four positions reduces RMSD to the level of hC3, as explained in the analysis of the two "transgenic" sequences (below). Thus, the displacement of sector II is mainly due to the change in its chemical composition. In the simulations of the rat complex, the displacement of this sector was even larger (RMSD =1.51-2.13 Å), possibly because the rat protein contains two more mutations (at positions 389 and 394 [31]). In contrast, sector I remained near the ligand (RMSD = 0.82-1.28 Å [31]).

The displacement of sector II in the rat complex [31] caused a loss or weakening of hydrogen-bonding and non-polar interactions between the protein and the ligand, in accordance with the lack of compstatin activity against rat C3. A similar loss of protein-ligand interactions is observed in the mouse complex, as detailed below. From a structural perspective, this is also reflected in the larger RMSD values of the ligand conformation in the mouse complex (next-to-last column of Table I). Alignment with respect to the experimental structure (last column) shows that the ligand preserves the shape of the bound conformation seen in the human complex [4], especially in runs M1-M2. Thus, these larger RMSD values are not due to structural distortions of the ligand, but mainly due to ligand

displacements away from the bound position. Indeed, the ligand maintains its secondary structure in all mouse and "transgenic" complexes (not shown), with the exception of an intramolecular β-bridge 3-11, which is not well preserved in the mouse runs (Supporting Information Fig. SF1).

**Protein-ligand interactions—**Statistics of intermolecular and intramolecular (ligand) hydrogen-bonding pairs in the mouse simulations is summarized in Table II. Figs. 3C-D plot intermolecular interaction energies for W4A9 and C3c residues. Some of the key stabilizing interactions in the human complex retain their strength in the mouse complex: The Trp4 mainchain amide maintains a strong interaction with the Arg456 sidechain (Fig. 2C and Table II). The Trp7 sidechain is positioned between sectors 455–458 and 488–491, making a very stable hydrogen bond with Met457 CO. The Ala9 amide forms a very stable hydrogen bond with the Asp491 side chain, and the His10 sidechain forms a somewhat weaker hydrogen bond with Asp491 (Fig. 2C and Table II). In general though, most protein-ligand interactions become weaker in the mouse complex: First, most intermolecular hydrogen bonds have smaller average occupancies in the mouse complex (Table II). Furthermore, several protein-ligand side-chain contacts are reduced, as shown in the side-chain contact maps of Figs. SF2-SF3. In particular, the mutation Pro393Asn (Fig. 1) eliminates a nonpolar contact with the Trp4 sidechain, which adopts a different orientation with respect to the human complex (Figs. 3B, 3D); the mutation Leu454His eliminates non-polar contacts with Val3 and His10. The reduction of intermolecular interaction energies is also clearly shown in Figs. SF4-B, which plot interaction energy *differences* (human-mouse) for the residues of Fig. 3. The largest changes involve Asn390 (loses its hydrogen-bonding interaction with Cys2 – see Table II), Asp491 (loses a hydrogen-bonding interactions with the His10 mainchain) and Arg456 (loses polar interactions with Trp4 and non polar interactions with residue Gln5). From the point of view of the ligand, the N-terminal moiety residues 1-5 and His10 show an approximately uniform reduction in energies (~3 kcal/mol). This destabilization of the mouse complex is caused by the displacement of sectors 388-393 and 344-349 (Fig. 2C and Table I), and is consistent with the experimental lack of compstatin activity against mouse C3.

From this analysis, it follows that the insertion of human-like mutations in sectors I-III might restore lost or diminished interactions with the ligand (residue moiety 1-7) and enhance the "transgenic"-protein affinity for compstatin. If these mutations stabilize the structure of the complex, they may also improve the key interactions of Asp491 (in sector IV) with Ala9 and His10. The effects of these mutations are discussed in the following section.

## The "Transgenic" complexes

We simulated two complexes (mhaC3:W4A9 and mhbC3:W4A9) of the mouse protein, respectively with four and six residue substitutions, taken from the human C3 sequence (see *Methods*). The first complex had six substitutions in sectors II and III (Pro392His, Asn393Pro, Arg395Gln, Gln396Lys, His454Leu, Pro459Arg); the second had three more mutations in sectors I and III (Ala345Gly, Thr457Met and Gly460Ala). The reasoning was explained in the *Methods* and is further detailed in the *Discussion* section. We conducted two 7-ns simulations (MHA1-MHA2) with the first complex, and four 8.75-ns simulations (MHB1-MHB4) with the second complex.

**Simulation conformations—**For both "transgenic" complexes, the protein backbone RMSD from the crystallographic conformation is comparable to the value computed for the mouse and human complexes (second column of Table I), showing that the introduction of human residues does not perturb significantly the total structure. Interestingly, the

conformational deviations of protein sectors I and II are reduced to the level of the human complex. This is also seen in Figs. 2E-F, which show the conformation of the compstatin binding site at the end of "transgenic" sequence run MHB4. The displacement of the ligand conformation is also significantly reduced, even though it remains slightly larger with respect to the simulations of the human complex. In addition, the 3-11 β-bridge is accurately preserved in the "transgenic" complex (Supporting Information Fig. SF1), as in the human complex [31]. A trajectory of the run MHB4 is shown in Supporting Information Video MHB4.

The somewhat larger RMSD values of sector IV, encountered in runs MHB1-3 of the second "transgenic" sequence, are due to transitions in the main-chain torsional angles of residues Pro488 and Gly489. These transitions are not associated with any mutations in this sector, which is invariant across species. They do not have a significant impact in the protein-ligand interactions and affinity; for example, the key residues Asp491 and Leu492 have small RMSD values in all simulations and retain strong interactions with the ligand, as detailed below.

**Protein-ligand interactions—**Along with the reduction in structural deviations from the human complex, the protein-ligand hydrogen-bonding interactions are also significantly improved in the "transgenic" complexes (Table II). Hydrogen-bond occupancies between the ligand and residues 455, 456 457 and 459 (sector III) or 491 (sector IV) are recovered, or even strengthened with respect to the human complex. Gly345 and Asn390 also improve their interactions, respectively, with Trp4 and the N-terminal moiety of the ligand, especially in the case of the second "transgenic" sequence. Important non-polar contacts, involving ligand-protein pairs Val3 – Leu454 and Trp4 – Asn390/His392/Pro393 in the human complex [4,31], are also recovered (Supporting Information Figs. SF2,SF3).

Figs. 3E-F display protein and ligand residue interaction energies for the second "transgenic" complex, mhbC3:W4A9. The strongest interactions involve protein residues Asp491 of sector IV, Arg456, Met457 and Arg459 of sector III, Asn390 of sector II and Gly345 of sector I. Asp491 makes two relatively strong hydrogen bonds with His10 (with occupancies ~65%); Arg456 forms a strong hydrogen bond with the main chain of residue Trp4 and a non-polar contact with the side chain of residue Gln5. The human - "transgenic" difference interaction energies, shown in Figs. SF4C-D, are consistently smaller (in absolute value) with respect to the human-mouse values of Figs. SF4A-B, reflecting the fact that the protein-ligand interactions of the human complex are better reproduced in the "transgenic" complex. The negative difference energies of residues 388-390 indicate that the interactions between sector I and N-terminal residues 1-2 are not entirely recovered, despite the four human-residue substitutions at this location (Pro392His, Asn393Pro, Arg395Gln, Gln396Lys). On the other hand, the four substitutions in sector III (His454Leu, Thr457Met, Pro459Arg, Gly460Ala) increase some intermolecular interactions at the level of the human complex, or even higher. In particular, the Arg456 sidechain makes a stronger interaction with the mainchain amide of Trp4, and Asp6 makes an occasional interaction with Arg459. This overall stabilization assists interactions with between Asp491 of sector IV and the sidechain of His10 (Table II).

Residue interaction free-energy differences (mhb-mha) between the first and second "transgenic" complex are shown in Figs. SF4E-F; corresponding difference contact maps (mhb-mha) are reported in Supporting Information Fig. SF3. The protein-ligand interactions are in general stronger in the second "transgenic" sequence. The Thr457Met and Gly460Ala mutations do not seem to contribute to the improved affinity. On the other hand, the Trp4-Ala345 hydrogen-bonding interaction is increased by ~1 kcal/mol in MHB, and presumably contributes also to the strengthening of the Trp4-Arg456 hydrogen bond (Table II).

Furthermore, the interactions between Ala9/His10 of the ligand and Asp491 in sector IV are increased by ~2 kcal/mol. This behavior (a correlation between ligand affinity and the ligand-Asp491 interactions) was also observed in the human complex [31]. Interestingly, hydrogen-bonding interactions between Asp6 and Arg459 are also more frequently observed in MHB (Table II) and contribute partly to the stronger polar interaction energy of this complex.

## Comparison of binding affinities

The association free energies of the mouse and "transgenic" complexes, computed by Eqs. (1-2) in the MM-GBSA approximation [40], are reported in Table III. The table also lists the corresponding values for the human-complex runs [31].

The second column lists free energy estimates, under the assumption that the protein and ligand have identical conformations in the complex and their dissociated (free) states in solution. The W4A9 affinity for the mouse protein (−46.0 kcal/mol) is the same as the affinity for rat C3 [31], and weaker than the affinity for human C3 by 8.2 kcal/mol, in line with the experimental lack of compstatin activity against mouse C3 [30]. The weaker affinity is associated with reduced protein-ligand interactions, reflected by the less negative non-polar free-energy component (−51.2 kcal/mol, compared to −60.2 kcal/mol in the human complex) and the less negative polar interaction energy (−31.5 kcal/mol, compared to −47.4 kcal/mol). Despite the weaker polar interactions, the penalty for burial of polar groups upon association is also smaller in the mouse complex, yielding comparable polar free energy components (+5.1 and +6.1 kcal/mol in the two complexes). The two "transgenic" complexes have association free energies of −51.1 kcal/mol and −54.3 kcal/mol, closer to the human complex. The second "transgenic" complex forms improved polar interactions (in line with the somewhat stronger protein-ligand hydrogen-bonding occupancies of Table II). At the same time, the non-polar interactions of both "transgenic" complexes are slightly worse with respect to the human complex.

The neglect of water-mediated interactions, that is inherent in the MM-GBSA approximation, may affect the accuracy of the computed affinities. Such interactions are not frequently observed in the present simulations and should not affect significantly the results. The most important interaction involves pair Trp4-Asn390; it is observed in ~30 % of the MHB snapshots and less than 13 % of the MD snapshots in the other complexes.

The association free energies in the last column of Table III contain contributions from protein and ligand structural relaxation. To compute these values, we conducted additional runs of the free proteins and ligand in solution, using the same protocol as in the complexes. The free-energy penalty due to the structural relaxation varies among complexes, but the qualitative conclusions still stand: the "transgenic" C3 affinities for W4A9 are similar to the affinity of human C3, and stronger than the affinity of mouse C3. The uncertainties (in parentheses) increase with respect to the values without structural relaxation (parentheses in column 2). Fig. SF5 plots running averages of the affinity free energies without structural relaxation (column 2); Fig. SF6 plots running averages for the total free energies of the complexes, free proteins and free ligand. Individual runs sample different conformations and converge to different limiting values by the end of the simulations, contributing to the resulting uncertainties. In the case of structural relaxation, the uncertainties are larger, due to contributions from the independent free protein and free ligand runs.

End-point methods such as MM-GBSA and the related Molecular Mechanics/Poisson Boltzmann Surface Area (MM-PBSA) approximation [40] have been used extensively to compute affinities for protein-ligand complexes (Ref. [44] and references therein) and their connection with statistical thermodynamics has been outlined in Ref. [45]. They are based

on several assumptions: they combine a molecular mechanics energy function with an implicit treatment of solvation effects and include solute conformational entropy effects in an approximate manner (Ref. [43] for a recent criticism and references therein). Here, the obtained relative ("transgenic"-mouse) association free-energies are ~ −6 – −9 kcal/mol (without relaxation) and ~ −9 – −10 kcal/mol (with relaxation). Even though the corresponding uncertainties are significant (especially with relaxation), the observed trends suggest that the inserted mutations have conferred approximately human-like affinity for W4A9 in mouse C3. These conclusions need experimental verification, which is out of the scope of the present article.

## 4. Discussion

In the present study we have examined by MD simulations a set of "transgenic" variants of mouse C3 with promising affinity for compstatin, an established inhibitor of human C3. We derived these variants by combining the mouse C3 sequence with 6-9 residue substitutions from the human C3 sequence. The substitutions were guided by the structure of the human C3:W4A9 complex [4] and insights on the structure and interactions of complexes between W4A9 and the human, rat or mouse C3 (Ref. [31] and this work).

The obtained "transgenic" variants have significantly improved affinities for W4A9, compared to the parent mouse and the related rat C3 protein [31], and are estimated to bind W4A9 as strongly as human C3. These predictions need to be tested experimentally. Nevertheless, they are supported by the consistency that characterizes the behavior of our MD simulations with complexes from five different C3 "species" (human and rat [31], mouse, and two "transgenic" variants of C3): The simulations of the human complex reproduce very well the crystallographic structure and interactions [4,31]; those of the rat and mouse complexes exhibit similar, localized conformational changes in the vicinity of the ligand, linked with a concomitant reduction in ligand affinity. The conformational changes are reproducible in several runs of rat, mouse and "transgenic" complexes with W4A9 (this work and [31]), as well as over twenty simulations of rat/mouse C3 complexes with a large number of other compstatin analogs (Tamamis et al, unpublished). In the "transgenic" complexes the conformational changes are reduced progressively as the similarity of the compstatin binding site with its human counterpart increases. Furthermore, the simulations predict a strengthening of binding affinity (from mouse/rat to human), in agreement with the experimentally determined species-specificity of compstatin [8,30].

The reasoning guiding the chosen mutations was explained in the *Methods* section. One question is whether a smaller subset of mutations could bring the protein affinity to the level of human C3. The simulations of the rat and mouse complexes show that residues 388-396 move away from the ligand, resulting in the loss of interactions with its N-terminal moiety and in the overall destabilization of the binding location/orientation of the ligand in the complex. A particularly noticeable difference among primate and non-primate residues in sector II involves the shift of a Pro residue from position 393 (primate) to 392 (non-primate). Test simulations showed that the introduction of only two human substitutions at positions 392-393 is not sufficient to retain the segment 388-396 near its crystallographic position. For this reason, four mutations were inserted in this region.

Two more key mutations emerged from the rat simulations [31]. Residue Leu454 of human C3 is converted to histidine in non-primate mammals, disrupting a hydrophobic cluster comprising Met346, Pro347 and Val3 of the ligand. Furthermore, Arg459 becomes proline in lower mammals, reducing non-polar interactions with Trp7. Reinstating the human residues Leu454 and Arg459 gives a total of six mutations for "transgenic" sequence mha. Analysis of the two runs of the mhaC3 complex showed that the sector 388-393 retained its

conformation, but some intermolecular (predominantly polar) interactions with Ala345, Arg456, Arg459 and Asp491 were not recovered. Among these residues, 345 varies between the first “transgenic” sequence (alanine) and the human protein (glycine). Even though the substitution Gly345Ala (from human to rat or mouse C3) does not weaken significantly its hydrogen bond with Trp 4 (Table II), the change in torsional angles may also hinder the formation of a strong hydrogen-bonding interaction between Trp4 and the side-chain of Arg456 (Table II). We thus designed a more extended “transgenic” sequence, containing the six substitutions of the fist sequence, Ala345Gly, and the two mutations Thr457Met and Gly460Ala, which render the sector 454-462 more human-like. The resulting protein reproduced quite well the structure, interactions and affinity of the human complex and represents a promising inhibitory target for compstatin in the mouse.

In conclusion, the MD-based strategy implemented in the current work suggests a way to design a compstatin-sensitive animal model. Given the species-specificity of known compstatin family peptides, such model will be useful for *in vivo* studies to efficiently assess suitability of new analogs to proceed to clinical trials. A similar strategy, involving a comparative analysis of MD simulations with a series of C3-peptide complexes is in progress to design novel “transgenic” sequences of compstatin family peptides. Overall, the paradigm of compstatin family peptides and their interactions with C3 in this and our previous works [14,15,29,31] demonstrates that MD-based methods can be useful computational aids for the design of peptides and proteins with tailored physicochemical and geometric properties and targeted design of biological activities.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The simulations were performed on *Linux* clusters of the Biophysics group at the University of Cyprus. Some simulations were performed at an *IBM* cluster at the Cyprus Institute. PT acknowledges support from a senior Fulbright Scholarship by the Cyprus Fulbright Commission. GA and PT acknowledge support from a University of Cyprus grant. CAF acknowledges financial support from the National Institutes of Health 5R01GM052032. DM acknowledges financial support from a UCR Bioengineering Center seed fund. GA and PT would like to thank Prof. Roland Dunbrack for the *SCWRL4* program. The authors acknowledge useful discussions with Aliana López de Victoria, Ron D. Gorham Jr., Chris A. Kieslich, and Noriko Sausman.

## References

1. Muller-Eberhard, HJ. Complement-Chemistry and pathways. In: Gallin, JI.; Goldstein, IM.; Synderman, R., editors. Inflammation: Basic principles and clinical correlates. 2nd Edition. Raven Press; New York: 2002. p. 21-54.
2. Anderson DH, Radeke MJ, Gallo NB, Chapin EA, Johnson PT, Curletti CR, Hancox LS, Hu J, Ebright JN, Malek G, Hauser MA, Rickman CB, Bok D, Hageman GS, Johnson LV. The pivotal role of the complement system in aging and age-related macular degeneration: hypothesis re-visited. Prog. Retin. Eye. Res. 2010; 29(2):95–112. [PubMed: 19961953]
3. Sahu A, Lambris JD. Complement inhibitors: a resurgent concept in anti-inflammatory therapeutics. Immunopharmacology. 2000; 49:133–148. [PubMed: 10904113]
4. Janssen BJC, Halff EF, Lambris JD, Gros P. Structure of Compstatin in Complex with Complement Component C3c Reveals a New Mechanism of Complement Inhibition. J Biochem. 2007; 282:29241–29247.
5. Ricklin D, Lambris JD. Compstatin: A Complement Inhibitor on its Way to Clinical Application. Adv Exp Med Biol. 2008; 632:273–292. [PubMed: 19025129]

6. Morikis, D.; Lambris, JD. Structure, dynamics, activity and function of compstatin and design of more potent analogs. In: Morikis, D.; Lambris, JD., editors. Structural Biology of the Complement System. CRC Press/Taylor & Francis Group; Boca Raton, FL: 2005.
7. Soulika AM, Holl MCH, Sfyroera G, Sahu A, Lambris JD. Compstatin inhibits complement activation by binding to the β-chain of complement factor 3. Mol Immunol. 2006; 43:2023–2029. [PubMed: 16472861]
8. Sahu A, Ray BK, Lambris JD. Inhibition of human complement by a C3- binding peptide isolated from a phage-displayed random peptide library. J Immunol. 1996; 157:884–891. [PubMed: 8752942]
9. Morikis D, Assa-Munt N, Sahu A, Lambris JD. Solution structure of compstatin, a potent complement inhibitor. Protein Sci. 1998; 7:619–627. [PubMed: 9541394]
10. Klepeis JL, Floudas CA, Morikis D, Lambris JD. Predicting peptide structures using NMR data and deterministic global optimization. J Comput Chem. 1999; 20:1354–1370.
11. Sahu A, Soulika A, Morikis D, Spruce L, Moore WT, Lambris JD. Binding kinetics, structure-activity relationship, and biotransformation of the Complement inhibitor Compstatin. J Immunol. 2000; 165:2491–2499. [PubMed: 10946275]
12. Morikis D, Roy M, Sahu A, Troganis A, Jennings PA, Tsokos GC, Lambris JD. The structural basis of compstatin activity examined by structure-function-based design of peptide analogs and NMR. J Biol Chem. 2002; 277:14942–14953. [PubMed: 11847226]
13. Morikis D, Lambris JD. Structural aspects and design of low-molecular-mass complement inhibitors. Biochem Soc Trans. 2002; 30:1026–1036. [PubMed: 12440966]
14. Mallik B, Lambris JD, Morikis D. Conformational interconversion in compstatin probed with molecular dynamics simulations. Proteins: Struct Func and Bioinf. 2003; 52:130–141.
15. Tamamis P, Skourtis S, Morikis D, Lambris JD, Archontis G. Conformational analysis of compstatin analogues with molecular dynamics simulations in explicit water. J Mol Graph Mod. 2007; 26:571–580.
16. Klepeis J, Floudas CA, Morikis D, Tsokos CG, Argyropoulos E, Spruce L, Lambris JD. Integrated computational and experimental approach for lead optimization and design of compstatin variants with improved activity. J. Am. Chem. Soc. 2003; 125:8422–8423. [PubMed: 12848533]
17. Soulika AM, Morikis D, Sarrias MR, Roy M, Spruce LA, Sahu A, Lambris JD. Studies of structure-activity relations of Complement inhibitor Compstatin. J Immunol. 2003; 170:1881–1890. [PubMed: 12902490]
18. Morikis D, Soulika AM, Mallik B, Klepeis JL, Floudas CA, Lambris JD. Improvement of the anti-C3 activity of compstatin using rational and combinatorial approaches. Biochem Soc Trans. 2004; 32:28–32. [PubMed: 14748706]
19. Mallik B, Katragadda M, Spruce LA, Carafides C, Tsokos CG, Morikis D, Lambris JD. Design and NMR characterization of active analogues of compstatin containing non-natural amino acids. J Med Chem. 2005; 48:274–286. [PubMed: 15634022]
20. Mallik B, Morikis D. Development of a quasi-dynamic pharmacophore model for anti-complement peptide analogues. J Am Chem Soc. 2005; 127:10967–10976. [PubMed: 16076203]
21. Chiu TL, Mulakala C, Lambris JD, Kaznessis YN. Development of a New Pharmacophore Model That Discriminates Active Compstatin Analogs. Chem Biol Drug Des. 2008; 72:249–256. [PubMed: 18844671]
22. Morikis D, Lambris JD. Structural aspects and design of low-molecular-mass complement inhibitors. Biochem Soc Trans. 2002; 30:1026–1036. [PubMed: 12440966]
23. Klepeis JL, Floudas CA, Morikis D, Tsokos CG, Lambris JD. Design of peptide analogs with improved activity using a de novo protein design approach. Ind & Eng Chem Res. 2004; 43:3817–3826.
24. Holland MCH, Morikis D, Lambris JD. Synthetic small molecule complement inhibitors. Current Opinion in Investigational Drugs. 2004; 5:1164–1173. [PubMed: 15573867]
25. Morikis, D.; Floudas, CA.; Lambris, JD. Structure-based integrative computational and experimental approach for the optimization of drug design. In: Sunderam, VS.; van Albada, GD.; Sloot, PMA.; Dongarra, JJ., editors. ICCS 2005, Lecture Notes in Computer Science: Computational Science. Springer-Verlag; Berlin-Heidelberg: 2005. p. 680-688.

26. Morikis D, Mallik B, Zhang L. Biophysical and bioengineering methods for the study of the complement system at atomic resolution. WSEAS Transactions on Biology and Biomedicine. 2006; 6:408–413.
27. Bellows ML, Fung HK, Taylor MS, Floudas CA, De Victoria A López, Morikis D. New compstatin variants through two de novo protein design frameworks. Biophys J. 2010; 98:2337–2346. [PubMed: 20483343]
28. Qu H, Magotti P, Ricklin D, Wu EL, Kourtzelis I, Wu Y-Q, Kaznessis YN, Lambris JD. Novel analogues of the therapeutic complement inhibitor compstatin with significantly improved affinity and potency. Molecular Immunology. 2011 In Press.
29. de Victoria, A López; Gorham, RD., Jr; Bellows, ML.; Ling, J.; Lo, DD.; Floudas, CA.; Morikis, DA. New Generation of Potent Complement Inhibitors of the Compstatin Family. Chemical Biology & Drug Design. 2011 In Press.
30. Sahu A, Morikis D, Lambris JD. Compstatin, a peptide inhibitor of complement, exhibits species-specific binding to complement component C3. Mol Immunol. 2003; 39:557–566. [PubMed: 12431389]
31. Tamamis P, Morikis D, Floudas CA, Archontis G. Species specificity of the complement inhibitor compstatin investigated by all-atom molecular dynamics simulations. Proteins: Struct. Func. and Bioinf. 2010; 5278:2655–2667.
32. Davis ME, Madura JD, Luty BA, McCammon JA. Electrostatics and Diffusion of Molecules in Solution: Simulations with the University of Houston Brownian Dynamics Program. Comput Phys Commun. 1991; 62:187–197.
33. Mackerell AD, et al. An all-atom empirical potential for molecular modelling and dynamics study of proteins. J. Phys. Chem. B. 1998; 102:3586–3616.
34. Mackerell AD Jr, Feig M, Brooks CL 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 Comput Chem. 2004; 25:1400–1415. [PubMed: 15185334]
35. Macias AT, Mackerell AD Jr. CH/pi interactions involving aromatic amino acids: Refinement of the CHARMM tryptophan force field. J Comput Chem. 2005; 26:1452–1463. [PubMed: 16088926]
36. Brooks BR, Brooks CL III, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. J Comput Chem. 2009; 30:1545–1614. [PubMed: 19444816]
37. Eswar, N.; Marti-Renom, MA.; Webb, B.; Madhusudhan, MS.; Eramian, D.; Shen, M.; Pieper, U.; Sali, A. Current Protocols in Bioinformatics. John Wiley & Sons, Inc; 2006. Comparative Protein Structure Modeling With MODELLER; p. 5.6.1-5.6.30.
38. Krivov GG, Shapovalov MV, Dunbrack RL Jr. Improved prediction of protein side-chain conformations with SCWRL4. Proteins: Struct Func and Bioinf. 2009; 77:778–795.
39. Seeber M, Cecchini M, Rao F, Settanni G, Caflisch A. Wordom: a program for efficient analysis of molecular dynamics simulations. Bioinformatics. 2007; 23(19):2625–2627. [PubMed: 17717034]
40. Massova I, Kollman PA. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. Perspect Drug Discov Des. 2000; 18:113–135.
41. Im W, Lee MS, Brooks CL III. Generalized Born Model with a Simple Smoothing Function. J Comput Chem. 2003; 24:1691–1702. [PubMed: 12964188]
42. Chen J, Im W, Brooks CL III. Balancing Solvation and Intramolecular Interactions: Toward a Consistent Generalized Born Force Field. J Am Chem Soc. 2006; 128:3728–3736. [PubMed: 16536547]
43. Singh N, Warshel A. Absolute binding free energy calculations: On the accuracy of computational scoring of protein–ligand interactions. Proteins: Struct Func and Bioinf. 2010; 78:1705–1723.
44. Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct. 2007; 36:21–42. [PubMed: 17201676]

45. Swanson JMJ, Henchmann R, McCammon JA. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. Biophys J. 2004; 86:67–74. [PubMed: 14695250]

46. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. ClustalW2 and ClustalX version 2. Bioinformatics. 2007; 23(21):2947–2948. [PubMed: 17846036]

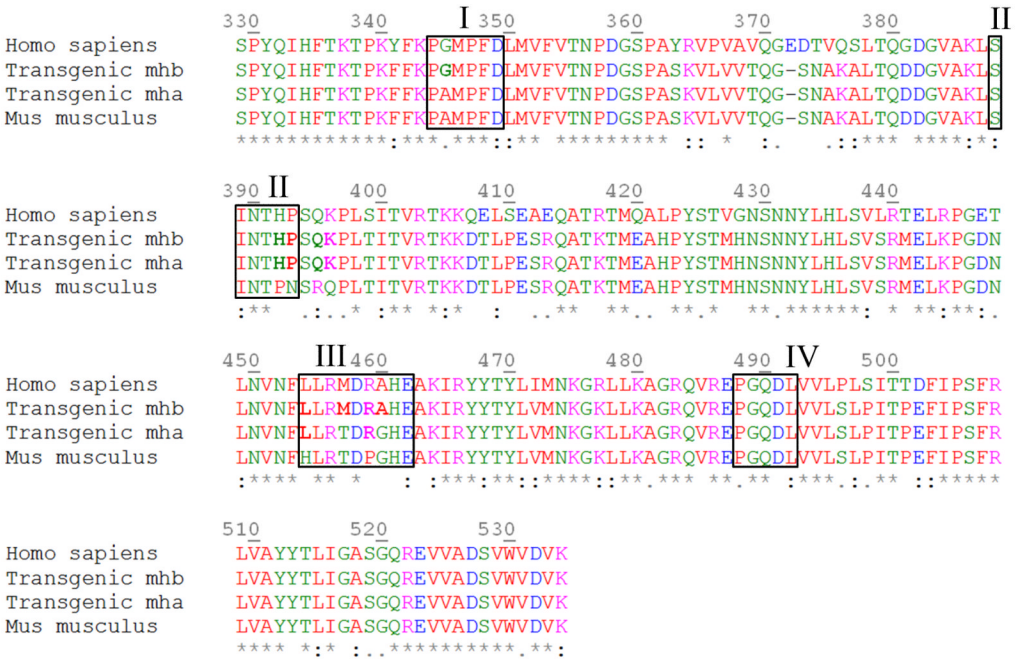


**Fig. 1.**

Alignment of human C3, mouse C3 and the two “transgenic” sequences of this work (mha and mhb). Only the region 329-534 of the human C3 compstatin-binding site is shown. The alignment was prepared with the program *CLUSTALW* v. 2.0.12 [46]. The color code used is: red - hydrophobic, green - polar, blue – negatively charged and purple – positively charged aminoacids. An asterisk (*) denotes invariant amino acids, a colon (:) strongly similar, and a period (.) weakly similar amino acids. The regions interacting with compstatin enclose in brackets. The substitutions contained in the “transgenic” sequences are shown in bold.

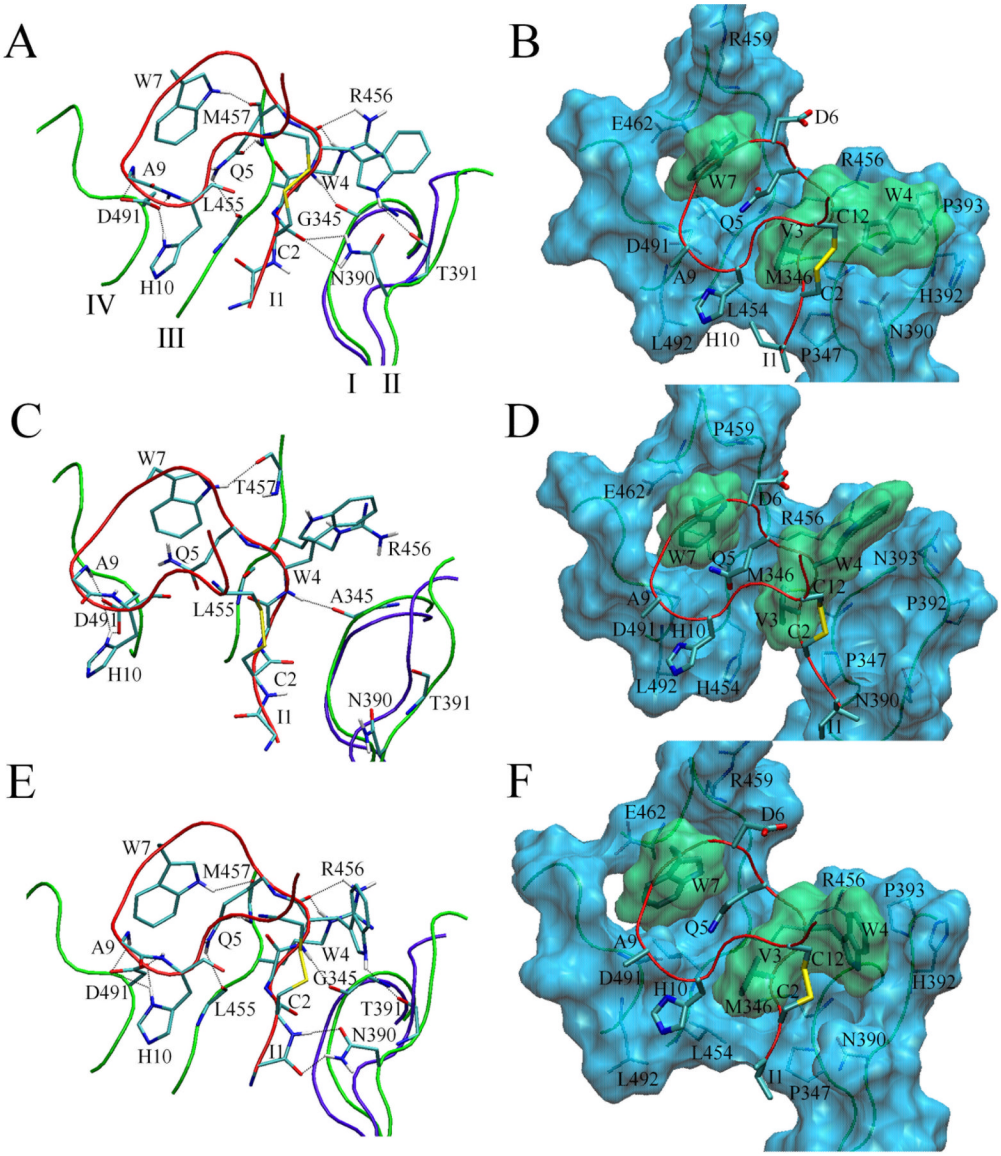


**Fig. 2.**

Simulation structures of the compstatin binding site for the human (A, B), mouse (C, D) and "transgenic" MHB complexes (E, F), respectively at the end of runs H2, M3 and MHB4. Important hydrogen bonds are shown in the left panels and non-polar contacts in the right panels. The labels I-IV (in A) indicate the four protein sectors with atoms at least within 7 Å from the ligand. Compstatin is shown in red tubes and sticks. The green tubes show the sectors I-IV. The violet tubes in A, C, E show the initial conformations of sectors I and II. The blue lines in A, C, E denote important hydrogen bonds. In B, D, F protein residues are additionally represented by a cyan color surface, and ligand residues Cys2, Val3, Trp4, Trp7 and Cys12 are additionally represented by a red color surface.

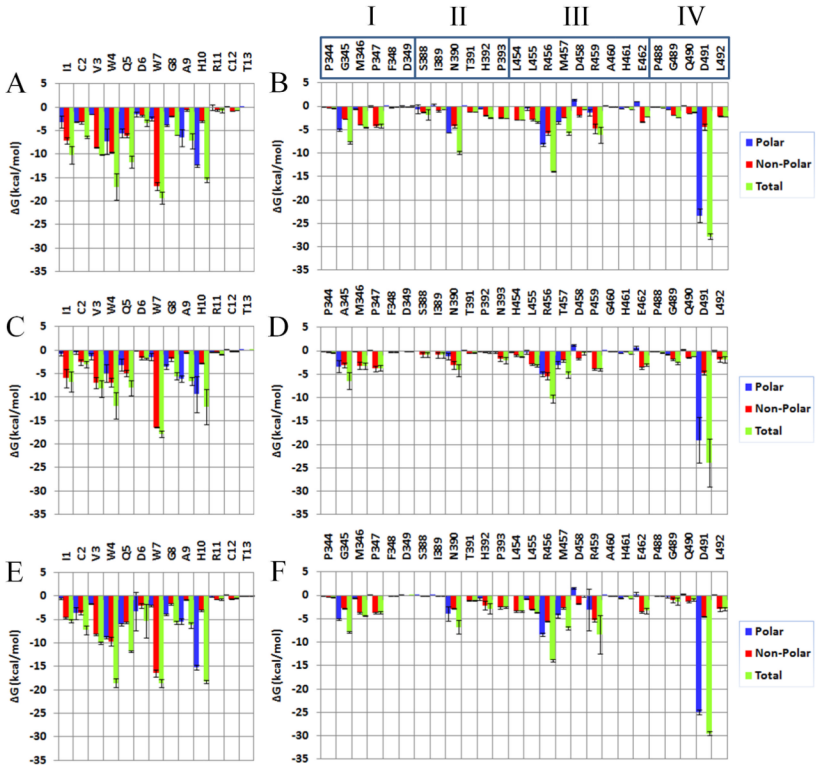


**Fig. 3.**

Intermolecular interaction free energies for selected ligand (left panels) and protein residues (right panels), computed by Eq. (3). **Top panel (A, B):** Human complex (from Ref. [31]); **Middle panel (C, D):** Mouse complex; **Lower panel (E, F):** Second “transgenic” complex (MHB). For each complex, the values are averaged over all corresponding runs. The uncertainties (error bars) are computed from the standard deviation of the average values.

**Table I**

Root mean square difference (RMSD) between the simulation coordinates of main-chain heavy atoms (N, $C_\alpha$, C) and the corresponding coordinates in the crystallographic structure of the human C3:W4A9 complex [31]. All values are reported in Å.

| Runs[a] | C3[b] | 344-349[c] | 388-393[c] | 454-462[c] | 488-492[c] | Compstatin | Compstatin[d] |
|---|---|---|---|---|---|---|---|
| H1 | 0.90 (0.90) | 0.93 (0.95) | 1.04 (0.90) | 1.15 (1.09) | 1.09 (1.23) | 1.73 (1.57) | 1.01 (0.85) |
| H2 | 0.71 (0.70) | 0.86 (0.77) | 0.99 (0.97) | 0.80 (0.87) | 0.64 (0.72) | 1.80 (1.60) | 1.20 (0.90) |
| M1 | 0.85 (0.86) | 1.55 (1.45) | 1.47 (1.45) | 1.06 (1.16) | 0.88 (1.05) | 2.89 (2.72) | 1.28 (1.21) |
| M2 | 0.86 (0.86) | 1.45 (1.05) | 1.67 (1.67) | 1.20 (1.19) | 0.84 (1.06) | 2.42 (2.14) | 0.90 (1.00) |
| M3 | 0.86 (0.91) | 1.24 (1.27) | 1.80 (1.93) | 1.29 (1.19) | 0.86 (1.35) | 3.43 (2.85) | 1.97 (1.24) |
| MHA1 | 0.81 (0.85) | 0.94 (1.21) | 1.22 (1.24) | 1.06 (1.14) | 1.07 (0.96) | 2.33 (2.75) | 1.38 (1.46) |
| MHA2 | 0.75 (0.78) | 1.27 (1.32) | 1.12 (1.28) | 0.90 (0.87) | 0.74 (0.78) | 2.01 (2.05) | 1.29 (1.16) |
| MHB1 | 1.02 (1.00) | 0.86 (0.87) | 1.22 (1.24) | 1.10 (0.97) | 2.24 (2.26) | 2.25 (2.04) | 1.21 (1.19) |
| MHB2 | 0.89 (0.87) | 0.68 (0.79) | 0.84 (0.92) | 1.16 (1.15) | 1.87 (1.73) | 2.06 (2.03) | 1.22 (1.22) |
| MHB3 | 0.93 (0.93) | 0.74 (0.92) | 0.99 (1.05) | 1.41 (1.37) | 2.17 (1.97) | 2.11 (2.12) | 1.32 (1.28) |
| MHB4 | 0.81 (0.79) | 1.00 (0.85) | 1.21 (1.11) | 1.00 (0.90) | 1.00 (0.87) | 2.20 (1.60) | 1.13 (1.02) |

All values (except the last column) are computed without rotation/translation and are averaged over the last 1 ns of the runs. In parentheses are the corresponding average RMSD values over the last 7 ns.

[a] H1-H2 denote the runs of the human C3:W4A9 complex, analyzed in Ref. [31]. M1-M3 are the runs of the mouse complex (mC3:W4A9); MHA1-MHA2 and MHB1-MHB4 are, respectively, the runs of the first and second "transgenic" C3 complexes (see *Methods*).

[b] All protein main-chain heavy atoms, excluding an external, harmonically restrained protein shell at least 20 Å away from the ligand (see *Methods* and [31]).

[c] Residues in these four sectors contain atoms within 7 Å from the ligand.

[d] RMSD values after alignment of the ligand main-chain atoms N, $C_\alpha$, C with respect to the experimental bound conformation.

**Table II**

Occupancies of important intermolecular and intramolecular (ligand) hydrogen-bonding atom pairs.

| | | Human Runs[a] [31] | | | Mouse Runs | | | | MHA Runs | | | MHB Runs | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Compstatin** | **Protein** | **1** | **2** | **Average** | **1** | **2** | **3** | **Average** | **1** | **2** | **Average** | **1** | **2** | **3** | **4** | **Average** |
| Ile1 OY | Asn390 ND2 | <1 | 4 | 2 | 25 | 33 | 18 | 25 | 49 | 41 | 45 | 0 | 0 | 0 | 0 | 0 |
| Ile1 N | Asn390 OD1 | 57 | 6 | 32 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ile1 O | Asn390 ND2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | <1 | 52 | 14 |
| Cys2 N | Asn390 OD1 | 79 | 66 | 73 | 10 | 31 | 6 | 16 | 39 | 38 | 39 | 91 | 99 | 74 | 66 | 83 |
| Cys2 N | Asn390 ND2 | <1 | 2 | 1 | 1 | 1 | <1 | 1 | 6 | <1 | 3 | 4 | <1 | 2 | 18 | 6 |
| Cys2 O | Asn390 ND2 | 37 | 74 | 56 | 12 | 4 | < | 5 | 39 | 31 | 35 | 0 | 98 | 63 | 2 | 41 |
| Trp4 N | Gly/Ala345 O | 97 | 100 | 99 | 33 | 80 | 88 | 67 | 71 | 75 | 73 | 99 | 99 | 99 | 100 | 99 |
| Trp4 NE1 | Thr391 O | 35 | 29 | 32 | <1 | <1 | <1 | 0 | 2 | 23 | 13 | 25 | 35 | 22 | 32 | 29 |
| Trp4 NE1 | Asn390 OD | 5 | 29 | 17 | <1 | 1 | 0 | 0 | 1 | 19 | 10 | 0 | 0 | 13 | 0 | 3 |
| Trp4 O | Arg456 NE | 33 | 98 | 66 | 40 | 87 | 71 | 66 | 75 | 60 | 68 | 98 | 97 | 98 | 94 | 97 |
| Trp4 O | Arg456 NH* | 38 | 47 | 43 | 12 | 23 | 27 | 21 | 24 | 35 | 30 | 36 | 54 | 27 | 41 | 40 |
| Gln5 OE1 | Met/Thr457 N | 77 | 98 | 88 | 80 | 40 | 34 | 51 | 97 | 91 | 94 | 98 | 99 | 99 | 99 | 99 |
| Gln5 NE2 | Leu455 O | 47 | 89 | 68 | 81 | 57 | 32 | 57 | 80 | 92 | 86 | 85 | 90 | 81 | 92 | 87 |
| Asp6 OD | Arg459 NH | 2 | 3 | 3 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | <1 | <1 | 2 | 78 | 20 |
| Asp6 O | Arg459 NH | 29 | <1 | 15 | 0 | 0 | 0 | 0 | 13 | 9 | 11 | 11 | 16 | 9 | 35 | 18 |
| Trp7 NE1 | Met/Thr457 O | 100 | 100 | 100 | 99 | 99 | 94 | 97 | 99 | 100 | 100 | 99 | 99 | 100 | 100 | 100 |
| Ala9 N | Asp491 OD* | 100 | 99 | 100 | 98 | 95 | 80 | 91 | 99 | 97 | 98 | 98 | 98 | 98 | 94 | 97 |
| His10 N | Asp491 OD* | 97 | 99 | 98 | 84 | 71 | 39 | 65 | 98 | 99 | 99 | 97 | 100 | 99 | 98 | 99 |
| His10ND1 | Asp491 OD* | 61 | 60 | 61 | 97 | 59 | 39 | 65 | 51 | 90 | 71 | 100 | 100 | 99 | 100 | 100 |
| Compstatin | Compstatin | 1 | 2 | Average | 1 | 2 | 3 | Average | 1 | 2 | Average | 1 | 2 | 3 | 4 | Average |
| Val3 N | His10 O | 100 | 99 | 100 | 82 | 41 | 56 | 60 | 97 | 99 | 98 | 97 | 99 | 70 | 99 | 91 |
| Val3 O | Cys12 N | 95 | 98 | 97 | 74 | 87 | 42 | 68 | 96 | 90 | 93 | 96 | 97 | 87 | 100 | 95 |
| Val3 O | Gln5 N | 98 | 98 | 98 | 91 | 63 | 73 | 76 | 95 | 97 | 96 | 95 | 99 | 90 | 97 | 95 |
| Gln5 NE2 | Trp7 O | 58 | <1 | 32 | 11 | 48 | 63 | 41 | 8 | <1 | 4 | 8 | 7 | 4 | 7 | 7 |
| Trp7 NE1 | Gln5 OE | 49 | 78 | 64 | 67 | 38 | 34 | 46 | 85 | 71 | 78 | 85 | 84 | 79 | 87 | 84 |

Hydrogen-bond occupancies (%) have been computed from the analysis of 700 snapshots (per run), extracted from the last 7-ns of the simulations, at 10-ps intervals. A hydrogen bond was present if the donor (D) – acceptor (A) distance was less than 3.5 Å and the corresponding angle (D-H⋯A) was larger than 90°.

[a]The "Human runs" (1-2) were analyzed in Ref. [31]. "Mouse" (1-3) are the runs of the mouse complex; MHA (1-2) and MHB (1-4) are, respectively, the runs of the first and second "transgenic" C3 complexes.

**Table III**

Average association free energies of the human, mouse and "transgenic" complexes, computed from the MD simulations (all values in kcal/mol).

| Run [a] | Free energy of binding | | | | |
|---|---|---|---|---|---|
| | Without structural relaxation[b] | | | | With structural relaxation[c] |
| | Total | Polar Component | Non-polar Component | Polar Interaction | |
| M1 | −42.3 | 3.6 | −45.9 | −35.8 | −41 |
| M2 | −50.1 | 5.2 | −55.3 | −33.5 | −50 |
| M3 | −45.7 | 6.6 | −52.3 | −25.3 | −30 |
| Average | **−46.0** (3.2) | 5.1 (1.2) | −51.2 (3.9) | −31.5 (4.5) | **−40** (9) |
| MHA1 | −51.6 | 7.6 | −59.2 | −41.1 | −50 |
| MHA2 | −50.6 | 5.9 | −56.5 | −43.7 | −50 |
| Average | **−51.1** (0.5) | 6.7 (0.9) | −57.8 (1.4) | −42.4 (1.3) | **−50** (5) |
| MHB1 | −52.4 | 3.2 | −55.6 | −47.7 | −51 |
| MHB2 | −54.1 | 3.6 | −57.6 | −50.7 | −53 |
| MHB3 | −53.7 | 3.0 | −56.7 | −48.3 | −37 |
| MHB4 | −57.2 | 1.0 | −58.2 | −55.3 | −55 |
| Average | **−54.3** (1.8) | 2.7 (1.0) | −57.0 (0.9) | −50.5 (3.0) | **−49** (9) |
| H1[31] | −54.8 | 6.8 | −61.5 | −48.2 | −43 |
| H2[31] | −53.5 | 5.4 | −58.9 | −46.5 | −53 |
| Average | **−54.2**(0.7) | 6.1(0.7) | −60.2(1.3) | −47.4(0.9) | **−48** (6) |

[a] M1-M3 are the runs of the mouse complex; MHA1-MHA2 and MHB1-MHB4 are, respectively, the runs of the first and second "transgenic" C3 complex. H1-H2 are the runs of the human complex, from Ref. [31].

[b] The conformations of the free protein and ligand are taken from the simulations of the complex. Polar and non-polar components are defined in Eq. (2). The polar *interaction* components [Eq. (3)] measure the strength of intermolecular polar (Coulomb and GB) interactions in the complexes. The values in parentheses are the standard deviations of the averages over different runs.

[c] The conformations of the free protein and ligand are obtained from separate MD simulations (see *Methods*); the free energy is obtained from Eq. (1); individual free energies for the complexes ( $G_{PL}$ ), free proteins ( $G_P$ ) and free ligand ( $G_L$ ) are listed in table ST1 of SI. Uncertainties (in parentheses) are computed by error propagation of the corresponding uncertainties in the free energies of the complexes, free proteins and free ligand (see Table ST1).