Design of Peptide Analogues with Improved Activity Using a Novel de Novo Protein Design Approach

J. L. Klepeis,[†] C. A. Floudas,*[,†] D. Morikis,[‡] C. G. Tsokos,[§] and J. D. Lambris[§]

Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544-5263, Department of Chemical and Environmental Engineering, University of California at Riverside, Riverside, California 92093-0359, and Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Recent advances in the treatment of the peptide design problem have led to the ability to select novel sequences given the structure of a peptide backbone. Despite these breakthroughs, issues related to the stability and functionality of these designed peptides remain sources of frustration. In this work, a novel method that addresses these issues for the computational design of peptides and proteins is introduced. The method is based on (i) in silico sequence selection for a template fold using a novel integer linear program (ILP) formulation and (ii) validation of fold stability and specificity through rigorous calculations of ensemble probabilities for the selected sequences. Through experimentally based functional analysis, the approach is shown to provide several peptide sequences with 6−7-fold improvement in activity over the synthetic therapeutic peptide compstatin, a 13-residue cyclic peptide that binds to complement component C3 and inhibits complement activation.

1. Introduction

The problem of peptide and protein design, first suggested almost two decades ago, begins with a known three-dimensional protein structure and requires the determination of an amino acid sequence compatible with this structure. At the outset the problem was termed the "inverse folding problem"[1,2] given that protein design has intimate links to the well-known protein folding problem.[3] In contrast to the characteristic of protein folding to associate a given protein sequence with its own unique shape, the inverse folding problem exhibits high levels of degeneracy; that is, a large number of sequences will be compatible with a given protein structure, although the sequences will vary with respect to properties such as activity and stability.

Experimentalists have explored the realm of sequence space through the application of mutagenesis, rational design, and directed evolution techniques to the problem of protein design. Rational design relies on the use of simple rules deduced from known protein structures, and although this approach has been successful, the allowable sequence space is highly restricted.[4,5] The goal of directed evolution is to obtain incremental improvements in protein properties by iterating between mutagenesis and screening.[6,7] Although the major benefit of these experimental techniques lies in their ability to appraise and improve properties that can be screened correctly, they also entail potential drawbacks. For one, there is a restriction on the number of mutants that can be screened experimentally, typically in the range of 10^3-10^6 sequences.[8] This capacity-related limitation is further compounded by the fact that single random mutations (e.g., random point mutagenesis by error-prone PCR[7]) have low probabilities for improving a particular property, while simultaneous multiple mutations tend to dilute the possibilities for improvement. One remedy has been the coupling of directed evolution and limited site-directed mutagenesis techniques, which transforms the difficulty into identifying the positions that should be targeted.[9]

In contrast, computational protein design allows for the screening of overwhelmingly large sectors of sequence space, with this sequence diversity subsequently leading to the possibility of a much broader range of properties and degrees of functionality among the selected sequences. Allowing for all 20 possible amino acids at each position of a small 50-residue protein results in 20^{50} combinations, or more than 10^{65} possible sequences. From this astronomical number of sequences, the computational sequence-selection process aims at selecting those sequences that will be compatible with a given structure using efficient optimization of energy functions that model the molecular interactions.

In an effort to make the difficult nature of the energy modeling and combinatorial optimization manageable, the first attempts at computational protein design focused only on a subset of core residues and explored steric van der Waals based energy functions through exhaustive searches for compatible sequences.[10,11] Over time, the models have evolved to incorporate improved rotamer libraries in combination with detailed energy models and interaction potentials. Although the consideration of packing effects on structural specificity is sometimes sufficient, as shown through the design of compatible structures using backbone-dependent rotamer libraries with only van der Waals energy evaluations for a subset of hydrophobic residues,[12,13] there has been extensive research into the development of models including hydrogen-bonding, solvent, and electrostatic effects.[14−17] These functional additions to the design

* To whom correspondence should be addressed. Address: Prof. Christodoulos A. Floudas, Department of Chemical Engineering, Princeton University, Princeton, NJ 08544-5263. Tel.: (609) 258-4595. Fax: (609) 258-0211. E-mail: floudas@titan.princeton.edu.

[†] Princeton University.

[‡] University of California at Riverside.

[§] University of Pennsylvania.

10.1021/ie0340995 CCC: $27.50
Published on Web 03/25/2004

models are especially important for full sequence design since packing interactions no longer dominate for noncore residues (e.g., surface and intermediate residues). The incorporation of these additional noncore residues increases the potential for diversity and therefore enhances the probability for improving functionality when compared to the parent system. An additional complication is the need to account for changes in amino acid compositions and inherent propensities through the appropriate definition of a reference state.[15,18,19] Overall, there is no consensus between model parametrizations, and it is unclear which methods are more valid and suitable for generic protein design.

Once an energy function has been defined, sequence selection is accomplished through an optimization-based search designed to minimize the energy objective. Both stochastic and deterministic methods have been applied to the computational protein design problem. Stochastic approaches are appealing because their heuristic nature can be used to control termination, and both genetic algorithms[20] and Monte Carlo methods[19,21] have been applied to the protein design problem. However, these methods involve some element of chance and thus can lack consistency and reliability in locating the global minimum.[22] Deterministic methods, such as the dead-end-elimination algorithm,[23] offer the advantage of convergence to a consistent solution. Nevertheless, these methods might not be globally deterministic in that heuristic modifications must be applied to make convergence tractable for complex systems.[19,24] In particular, one restriction for the dead-end-elimination method is the requirement for a pairwise representation of the energy function. More recent methods attempt to avoid the problem of optimizing residue interactions by manipulation of the shapes of free energy landscapes.[25]

Several sequence-selection approaches have been tested and validated by experiment, thereby firmly establishing the feasibility of computational protein design. The first computational design of a full sequence to be experimentally characterized was the achievement of a stable zinc-finger fold ($\beta\beta\alpha$) using a combination of a backbone-dependent rotamer library with atomistic level modeling and a dead-end-elimination-based algorithm.[26] Despite these accomplishments, the development of a computational protein design technique to rigorously address the problems of fold stability and functional design remains a challenge. One important reason for this is the almost universal specification of a fixed backbone, which does not allow for the true flexibility that would afford more optimal sequences and more robust predictions of stability. Moreover, several models that attempt to incorporate backbone flexibility highlight a second difficulty, namely, inadequacies inherent in energy modeling.[21] Obviously, there is a need for empirically derived weighting factors, but the dependence on specific heuristics might also hinder the generic nature of these computational protein design methods. Such modeling-based assumptions also raise issues regarding the appropriateness of the optimization method and underscore the question of whether it is sufficient merely to identify the globally optimal sequence or, more likely, a subset of low-lying energy sequences. An even more difficult problem relevant to both flexibility and energy modeling is to correctly model the interactions that control the functionality and activity of the designed sequences.

In this work, a novel two-stage computational peptide and protein design method is presented not only to select and rank sequences for a particular fold but also to validate the stability and specificity of the fold for these selected sequences. The sequence-selection phase relies on a novel integer linear programming (ILP) model with several important constraint modifications that improve the tractability of the problem and enhance its deterministic convergence to the global minimum. In addition, a rank-ordered list of low-lying energy sequences is identified, along with the global minimum-energy sequence. Once such a subset of sequences has been identified, the fold validation stage is employed to verify the stabilities and specificities of the designed sequences through a deterministic global optimization approach that allows for backbone flexibility. The selection of the best designed sequences is based on rigorous quantification of energy-based probabilities.

The directed formulation can be seamlessly integrated into an overall predictive method for structure stability and specificity, and this method can be used to predict designed peptide analogues with enhanced immunological activities.[27] The underlying hypothesis is that improved stability and specificity can lead to improved activity when certain functionally important features are conserved. This computational peptide and protein design approach is applied to the problem of property improvement for a synthetic cyclic peptide that inhibits complement activation. The wild-type system, compstatin, a 13-residue peptide with a disulfide bridge, has been resolved structurally via NMR spectroscopy.[28] The application of the presented approach has led to the identification of sequences displaying substantial improvements in inhibition activity, as validated through immunological assays.[27]

2. Theory and Methods

2.1. In Silico Sequence Selection. To correctly select a sequence compatible with a given backbone template, an appropriate energy function must first be identified. Desirable properties of energy models for protein design include both accuracy and rapid evaluation. Moreover, the functions should not be overly sensitive to fixed backbone approximations. In certain cases, additional requirements, such as the pairwise decomposition of the potential for application of the dead-end-elimination algorithm,[23] might be necessary.

Instead of employing a detailed atomistic-level model, which requires the empirical reweighting of energetic terms, the proposed sequence-selection procedure is based on optimizing a pairwise *distance-dependent* interaction potential. Such a statistically based empirical energy function assigns energy values for interactions between amino acids in the protein on the basis of the α-carbon separation distance for each pair of amino acids. Such structure-based pairwise potentials are fast to evaluate, and have been used in fold recognition and fold prediction.[29] One advantage of this approach is that there is no need to derive empirical weights to account for individual residue propensities. Moreover, the possibility that such interaction potentials lack sensitivity to local atomic structure is addressed within the context of the overall two-stage approach. In fact, the coarser nature of the energy function in the in silico sequence-selection phase might prove beneficial in that it allows for an inherent flexibility to the backbone.

A number of different parametrizations for pairwise residue interaction potentials exist. The simplest approach is the development of a binary version of the model such that each contact between two amino acids is assigned according to the residues types and the requirement that a contact be defined as occurring when the separation between the side chains of two amino acids is less than 6.5 Å.[30] An improvement of this model is based on the incorporation of a distance dependence for the energy of each amino acid interaction. Specifically, the α-carbon distances are discretized into a set of 13 bins to create a finite number of interactions, the parameters of which were derived from a linear optimization formulated to favor native folds over optimized decoy structures.[31,32] The use of a distance-dependent potential allows for the implicit inclusion of side chains and the specificity of amino acids. The resulting potential, which involves 2730 parameters, was shown to provide higher *Z* scores than other potentials and place native folds lower in energy.[31,32] Recent work has resulted in improvements through the use of physical constraints and extension of the parametrization to include *β*-carbon interactions to better represent side-chain placement.[33]

The linearity of the resulting formulation based on this distance-dependent interaction potential is also an attractive characteristic of the in silico sequence-selection procedure. The development of the formulation can be understood by first describing the variable set over which the energy function is optimized. First, consider the set $i = 1, ..., n$, that defines the number of residue positions along the backbone. At each position i, there can be a set of mutations represented by $j\{i\} = 1, ..., m_i$, where, for the general case $m_i = 20$, $\forall i$. The equivalent sets $k \equiv i$ and $l \equiv j$ are defined, and $k > i$ is required to represent all unique pairwise interactions. With this in mind, the binary variables y_i^j and y_k^l can be introduced to indicate the possible mutations at a given position. That is, the y_i^j variable indicates the type of amino acid that is active at a position in the sequence by taking the value of 1 for that specification. Then, the formulation, for which the goal is to minimize the energy according to the parameters that multiply the binary variables, can be expressed as

$$\min_{y_i^j, y_k^l} = \min \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=i+1}^{n} \sum_{l=1}^{m_k} E_{ik}^{jl}(x_i,x_k) y_i^j y_k^l$$

subject to

$$\sum_{j=1}^{m_i} y_i^j = 1 \qquad \forall\, i$$

$$y_i^j, y_k^l = 0 - 1 \qquad \forall\, i, j, k, l \tag{1}$$

The parameters $E_{ik}^{jl}(x_i,x_k)$ depend on the distance between the α-carbons at the two backbone positions (x_i, x_k), as well as the type of amino acids at those positions. The composition constraints require that there be exactly one type of amino acid at each position. For the general case, the binary variables appear as bilinear combinations in the objective function. Fortunately, this objective can be reformulated as a strictly linear (integer linear programming) problem[34]

$$\min_{y_i^j, y_k^l} = \min \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=i+1}^{n} \sum_{l=1}^{m_k} E_{ik}^{jl}(x_i,x_k) w_{ik}^{jl} \tag{2}$$

subject to

$$\sum_{j=1}^{m_i} y_i^j = 1 \qquad \forall\, i$$

$$y_i^j + y_k^l - 1 \le w_{ik}^{jl} \le y_i^j \qquad \forall\, i, j, k, l$$

$$0 \le w_{ik}^{jl} \le y_k^l \qquad \forall\, i, j, k, l$$

$$y_i^j, y_k^l = 0 - 1 \qquad \forall\, i, j, k, l$$

This reformulation relies on the transformation of the bilinear combinations to a new set of linear variables, w_{ik}^{jl}, and the addition of the four sets of constraints serves to reproduce the characteristics of the original formulation. For example, for a given i, j, k, l combination, the four constraints require w_{ik}^{jl} to be 0 when either y_i^j or y_k^l is equal to 0 (or when both are equal to 0). If both y_i^j and y_k^l are equal to 1, then w_{ik}^{jl} is also enforced to be 1.

The solution of the integer linear programming (ILP) problem can be accomplished rigorously using branch-and-bound techniques,[34,35] making convergence to the global minimum-energy sequence consistent and reliable. Furthermore, the performance of the branch-and-bound algorithm is significantly enhanced through the introduction of reformulation linearization techniques (RLTs). Here, the basic strategy is to multiply appropriate constraints by bounded nonnegative factors (such as the reformulated variables) and introduce the products of the original variables by new variables in order to derive higher-dimensional lower-bounding linear programming (LP) relaxations for the original problem.[36] These LP relaxations are solved during the course of the overall branch-and-bound algorithm and, thus, speed convergence to the global minimum. The following set of constraints illustrates the application of the RLT approach to the original composition constraint, the first set of constraint equations listed in formulation 2. First, the equations are reformulated by forming the product of the equation with some binary variables or their complements. For example, by multiplying by the set of variables y_k^l produces the following additional set of constraints

$$y_k^l \sum_{j=1}^{m_i} y_i^j = y_k^l \qquad \forall\, i, k, l \tag{3}$$

Equation 3 can now be linearized using the same variable substitution as introduced for the objective. The set of RLT constraints then becomes

$$\sum_{j=1}^{m_i} w_{ik}^{jl} = y_k^l \qquad \forall\, i, k, l \tag{4}$$

Finally, for such an ILP problem it is straightforward to identify a rank-ordered list of the low-lying energy

sequences through the introduction of integer cuts[34] and repetitive solution of the ILP problem.

In general, these (ILP) formulations belong to the class of NP complete problems.[34] Most available (ILP) codes use a branch-and-bound procedure to search for an optimal integer solution by solving a sequence of related LP relaxations. By using the enhancements outlined above, in combination with the commercial (LP) solver CPLEX,[35] a globally optimal (ILP) solution is generated in less than 5 CPU min on an HP J-2240 workstation.

2.2. Fold Stability and Specificity. Once a set of low-lying energy sequences has been identified via the sequence-selection procedure, the fold stability and specificity validation stage is used to identify the most optimal sequences according to a rigorous quantification of conformational probabilities. The foundation of the approach is grounded on the development of conformational ensembles for the selected sequences under two sets of conditions. In the first circumstance, the structure is constrained to vary, with some imposed fluctuations, about the template structure. In the second condition, a free-folding calculation is performed for which only a limited number of restraints are likely to be incorporated (in the case of compstatin and its analogues, only the disulfide bridge constraint is enforced) and with the underlying template structure not being enforced. In terms of practical considerations, the distance constraints introduced for the template-constrained simulation can be based on the structural boundaries defined by the NMR ensemble (in the case of compstatin and its analogues, a deviation of 1.5 Å is allowed for each nonconsecutive Cα−Cα distance from the known NMR structures), or they can simply allow some deviation from a subset of distances provided by the structural template, hence allowing for a flexible template on the backbone.

The formulations for the folding calculations are reminiscent of structure prediction problems in protein folding.[37] In particular, a novel constrained global optimization problem first introduced for structure prediction using NMR data[38] and later employed in a generic framework for the structure prediction of proteins[39] is employed. The global minimization of a detailed atomistic energy force field E_{ff} is performed over the set of independent dihedral angles, ϕ, that can be used to describe any possible configuration of the system. The bounds on these variables are enforced by simple box constraints. Finally, a set of distance constraints, E_l^{dis}, $l = 1, ..., N$, that are nonconvex in the internal coordinate system, can be used to constrain the system. The formulation is represented by the following set of equations

$$\min_{\phi} E_{\text{ff}} \tag{5}$$

subject to

$$E_j^{\text{dis}}(\phi) \leq E_j^{\text{ref}} \; j = 1, ..., N$$

$$\phi_i^{\text{L}} \leq \phi_i \leq \phi_i^{\text{U}} \; i = 1, ..., N_\phi$$

Here, $i = 1, ..., N_\phi$ corresponds to the set of dihedral angles, ϕ_i, with ϕ_i^{L} and ϕ_i^{U} representing lower and upper bounds, respectively, on these dihedral angles. In general, the lower and upper bounds for these variables are set to $-\pi$ and π. E_j^{ref} are reference parameters for the distance constraints, which assume the form of a typical square-well potential for both the upper and lower distance violations. The set of constraints is completely general and can represent the full combination of distance constraints or smaller subsets of the defined restraints. The force-field energy function, E_{ff}, can take on a number of forms, although the current work employs the ECEPP/3 model.[40]

The folding formulation represents a general nonconvex constrained global optimization problem, a class of problems for which several solution methods have been developed. In this work, the formulations are solved via the αBB deterministic global optimization approach, a branch-and-bound method applicable to the identification of the global minimum of nonlinear optimization problems with twice-differentiable functions.[37,38,41−45] A converging sequence of upper and lower bounds is generated, with the upper bounds on the global minimum obtained by local minimizations of the original nonconvex problem and the lower bounds belonging to the set of solutions of the convex lower bounding problems that are constructed by augmenting the objective and constraint functions through the addition of separable quadratic terms.

In addition to identifying the global minimum-energy conformation, the global optimization algorithm provides the means for identifying a consistent ensemble of low-energy conformations.[44,46,47] Such ensembles are useful in deriving quantitative comparisons between the free-folding and template-constrained simulations. In this way, the complications inherent in the specification of an appropriate reference state are avoided because a relative probability is calculated for each sequence studied during this stage of the approach. The relative probability for template stability, p_{temp}, can be found by summing the statistical weights for those conformers from the free-folding simulation that resemble the template structure (denoted as the set temp), and dividing this sum by the total of the statistical weights for all conformers from the free-folding simulation (denoted as the set total).

$$p_{\text{temp}} = \frac{\sum_{i \in \text{temp}} \exp(-\beta E_i)}{\sum_{i \in \text{total}} \exp(-\beta E_i)} \tag{6}$$

Here, $\exp(-\beta E_i)$ is the statistical weight for conformer i. Figure 1 illustrates an intersection of the sets of conformers from the template and free-folding simulations. Specifically, the free energy (using a harmonic approximation to the entropy) is calculated for each unique conformer, and these values are used to calculate the corresponding statistical weight. The sum of all statistical weights represents the denominator of eq 6, whereas the sum in the numerator is computed only from those conformers that are structurally similar to the template. For compstatin, the template-constrained optimizations required approximately six CPU hours on a single P-III 600-MHz processor running Linux. The free-folding optimizations were run on a cluster of 64 P-III 600-MHz processors running Linux, and the parallelized branch-and-bound algorithm utilized about 4−5 h by wallclock per sequence.

2.3. Peptide Synthesis and Complement Inhibition Assays. Peptide synthesis and purification was performed as described previously.[48,49] The inhibitory

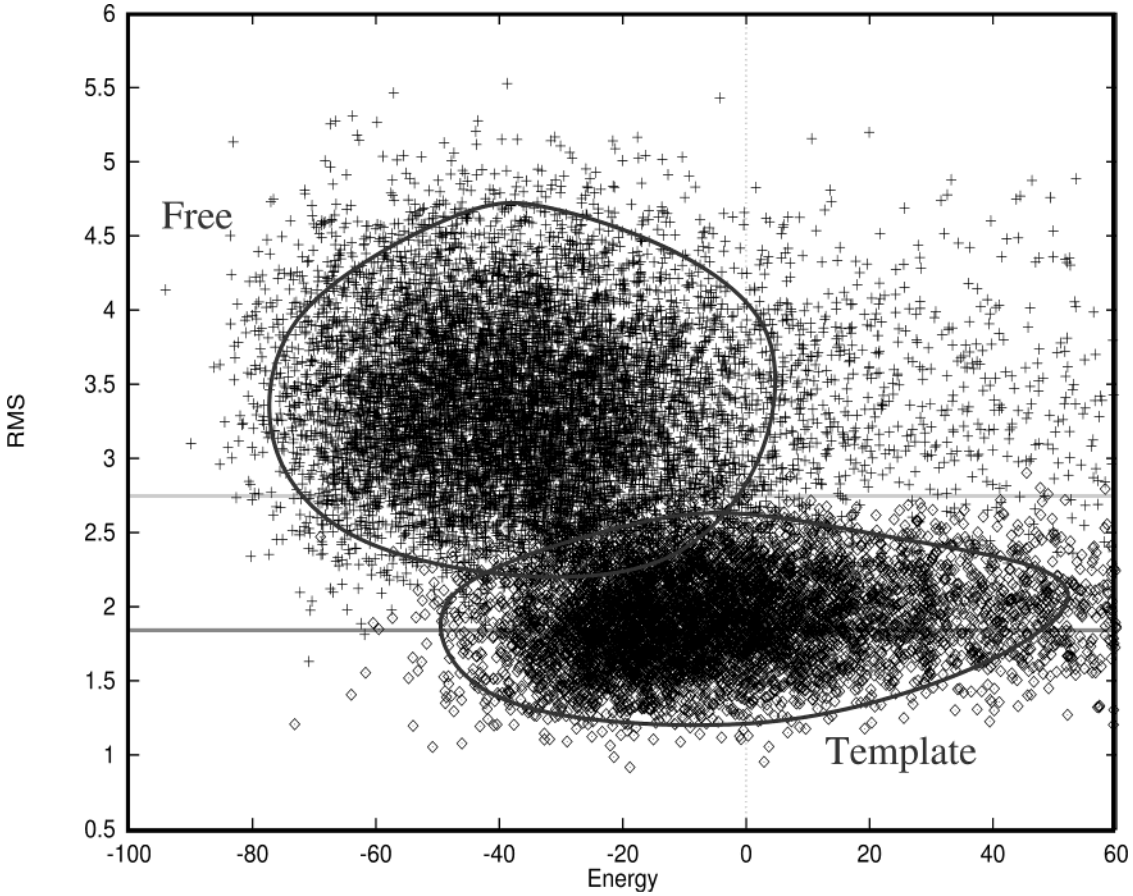

Figure 1. Illustration of intersection of sets, plotted as RMSDs (root-mean-squared deviations) from average template structure versus energy. The ellipses identify the general set of points from either the template or free-folding simulations. Horizontal lines indicate possible cutoff values for defining the allowed flexibility.

activities of compstatin and its analogues on the complement system were studied by measuring their effects on the classical pathway. Complement activation inhibition was assessed by measuring the inhibition of C3 fixation to OVA−anti-OVA complexes in normal human plasma. Briefly, microtiter plates were coated with ovalbumin, followed with anti-ovalbumin antibodies and normal human plasma (generally diluted 1/160) in the presence or absence of peptides diluted in gelatin Veronal buffer2+ (VBS, 0.1 gelatin, 0.5 mM $MgCl_2$, 2 mM $CaCl_2$). Complement activation was assessed using a goat anti-human C3 HRP conjugated antibody to detect deposition of activated C3b/iC3b. Color was developed by adding peroxidase substrate, and optical density was measured at 405 nm. The concentration of the peptide causing 50% inhibition of C3b/iC3b deposition was taken as IC50 and used to compare the activities of various peptides. All peptides were analyzed at least three times.

Results and Discussion

3.1. Compstatin: Wild-Type. Compstatin is a 13-residue cyclic peptide and a novel synthetic complement inhibitor with the prospect of being a candidate for development as an important therapeutic agent (see Figure 2). The binding and inhibition of complement component C3 by compstatin is significant because C3 plays a fundamental role in the activation of the classical, alternative, and lectin pathways of complement activation. Although complement activation is part of normal inflammatory response, inappropriate complement activation can cause host-cell damage, which is the case in more than 25 pathological conditions, including autoimmune diseases, stroke, heart attack, Alzheimer's disease, and burn injuries.[50] Compstatin was initially identified through screening of a phage-displayed random peptide library with C3b, a proteolytically activated form of complement C3, and later

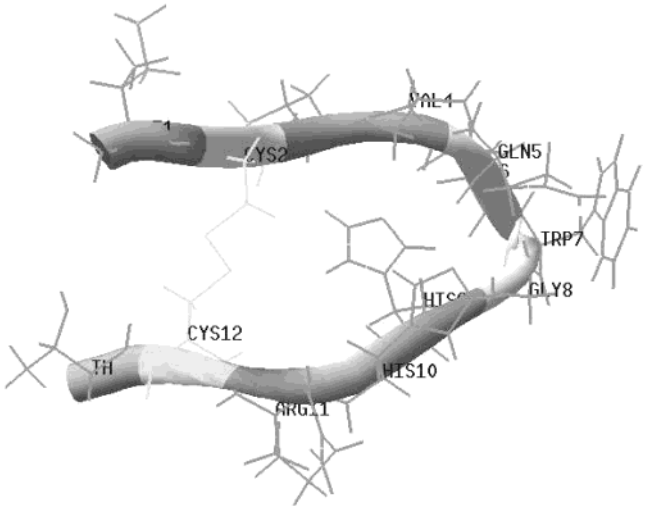

Figure 2. Sequence of compstatin in its smallest active form is Ile^1-Cys^2-Val^3-Val^4-Gln^5-Asp^6-Trp^7-Gly^8-His^9-His^{10}-Arg^{11}-Cys^{12}-Thr^{13}-NH_2, with Cys^2 and Cys^{12} forming a disulfide bond. The structure of compstatin was solved and shown to include a type I β-turn between residues Gln^5-Asp^6-Trp^7-Gly^8, with NOE data suggesting the formation of this flexible yet definitive type β-turn.[28]

truncated to a 13-residue peptide without loss of activity.[48] More recently, compstatin has been tested in a variety of clinically relevant models.

The sequence of compstatin in its smallest active form is Ile^1-Cys^2-Val^3-Val^4-Gln^5-Asp^6-Trp^7-Gly^8-His^9-His^{10}-Arg^{11}-Cys^{12}-Thr^{13}-NH_2, with Cys^2 and Cys^{12} forming a disulfide bond. The structure of compstatin was solved and shown to have a type I β-turn between residues Gln^5-Asp^6-Trp^7-Gly^8, with NOE data suggesting the formation of this flexible yet definitive type β-turn. It is significant to add that several immunogenic peptides adopt functional β-turn structures, which further support the promise of compstatin for future drug development.

As a result of its importance, compstatin has been the subject of extensive structure−function studies involving a variety of peptide analogues.[49,51] Along with the turn structure-spanning residue Gln^5-Gly^8, the termini of compstatin form a hydrophobic cluster stabilized by the disulfide bond and involving residues Ile^1, Cys^2, Val^3, Val^4, Cys^{12}, and Thr^{13}. Moreover, acetylation, in addition to preventing disruption of this hydrophobic cluster by removing the positive charge from the amino terminus, was found to significantly reduce proteolytic susceptibility and N-terminal processing of the peptide.[49]

It is well-known that turns are potential sites for molecular recognition, and initial observations had suggested that the turn residues of compstatin were important for inhibitory activity. However, it was unclear as to whether the β-turn was acting only to provide structural stability or whether it was also important for functional recognition. β-turn substitution analogues demonstrated that other sequences with type I turn preferences were functionally inactive and suggested that side-chain interactions exist between turn residues and C3.[49] In total, analogues of compstatin have indicated that Val^3 and the four residues of the β-turn are essential for retaining activity.[49]

More recent functional and NMR-based structure analyses of rationally designed compstatin analogues have unveiled several important observations concerning structure−activity correlations.[51] The rationally designed analogues involved either single or double alanine substitutions between Cys^2 and Cys^{12}, as well as single, double, and triple substitutions of the four turn residues. In regard to the β-turn, it was determined that the turn structure is a necessary but not a sufficient condition for activity and that flexibility of the turn is also essential for activity. Moreover, it was determined that Trp^7 is likely to be involved in a direct interaction with C3. Consideration of the hydrophobic cluster also indicated that, in addition to being held together by the disulfide bridge, this component of compstatin is involved in binding and activity with C3, but is not alone sufficient for activity. In addition, the difficulty in the optimization of the compstatin system has been demonstrated using both experimental combinatorial and rational design techniques.[51] Both studies led to the identification of an approximately 2-fold more active analogue.

3.2. In Silico Sequence Selection. The first stage of the design approach involves the selection of sequences compatible with the backbone template through the solution of the ILP problem. The formulation relies only on the α-carbon coordinates of the backbone residues, which were taken from the NMR-average solution structure of compstatin.[28]

A full computational design study from compstatin would result in a combinatorial search of $20^{13} \approx 8 \times 10^{16}$ sequences. However, in light of the results of the experimental studies of the rationally designed peptides, a directed, rather than full, set of computational design studies was performed. First, because the disulfide bridge was found to be essential for aiding in the formation of the hydrophobic cluster and prohibiting the termini from drifting apart, both residues Cys^2 and Cys^{12} were maintained. In addition, because the structure of the type I β-turn was not found to be a sufficient condition for activity, the turn residues were fixed to be those of the parent compstatin sequence, namely, Gln^5-Asp^6-Trp^7-Gly^8. In fact, when stronger type I β sequences were constructed, which was supported by NMR data indicating that these sequences provided higher β-turn populations than compstatin, these sequences resulted in lower or no activity.[51] Therefore, the further stabilization of the turn residues, which would likely be a consequence of the computational peptide design procedure, might not enhance compstatin activity. This is especially true for Trp^7, which was found to be a likely candidate for direct interaction with C3. For similar reasons, Val^3 was maintained throughout the computational experiments.

After the compstatin system had been designed to be consistent with those features found to be essential for compstatin activity, six residue positions were selected to be optimized. Of these six residues, positions 1, 4, and 13 have been shown to be structurally involved in the formation of a hydrophobic cluster involving residues at positions 1, 2, 3, 4, 12, and 13, a necessary but not sufficient component for compstatin binding and activity. The remaining residues, namely, those at positions 9−11, span the three positions between the turn residues and the C-terminal cystine. For the wild-type sequence, these positions are populated by positively charged residues, with a total charge of +2 coming from two histidine residues and one arginine residue.

On the basis of the structural and functional characteristics of those residues involved in the hydrophobic cluster, positions 1, 4, and 13 were allowed to select only from those residues defined as belonging to the hydrophobic set (A, F, I, L, M, V, Y). In addition, this set included threonine for position 13 to allow for the selection of the wild-type residue at this position. In positions 9−11, all residues were allowed, excluding cystine and tryptophan. A set of 50 low-lying energy sequences are listed in Table 1. The cutoff of 50 low-lying energy sequences is somewhat arbitrary, and a larger set of sequences can easily be generated and passed to the second stage of the approach. More generally, the variation at each position can be characterized by statistical analysis. Table 2 summarizes the preferred selection at each position according to the composition of the lowest-lying energy sequences.

The sequence-selection results exhibit several important and consistent features. First, position 10 is dominated by the selection of a histidine residue, a result that directly reinforces the composition of the wild-type compstatin sequence. In contrast, position 11 is found to have the largest variation in composition, with both polar, hydrophobic and charged residue being part of the set of optimal low-lying energy sequences. At position 9, a subset of the residues chosen for position

Table 1. Example Set of 50 Low-Lying Energy Sequences Predicted by the in Silico Sequence-Selection Method

iter	energy	1	4	9	10	11	13	iter	energy	1	4	9	10	11	13
1	−85.473	A	Y	A	H	T	V	26	−81.133	A	Y	T	H	V	V
2	−85.120	V	Y	A	H	V	A	27	−81.038	A	Y	T	H	F	V
3	−84.923	V	Y	A	H	A	V	28	−80.789	A	Y	T	H	D	V
4	−84.862	V	Y	A	H	T	A	29	−80.734	A	V	F	H	H	F
5	−84.711	A	Y	A	H	V	A	30	−80.734	V	A	F	H	H	F
6	−84.683	A	Y	A	H	A	V	31	−80.481	A	Y	F	H	T	V
7	−84.453	A	Y	A	H	T	A	32	−80.367	V	Y	T	H	H	A
8	−84.265	A	V	A	H	T	V	33	−80.260	V	Y	T	H	F	A
9	−84.241	A	Y	A	H	F	F	34	−80.225	V	Y	T	H	A	A
10	−83.934	A	Y	A	H	H	V	35	−80.181	V	A	F	H	H	V
11	−83.912	V	V	A	H	V	A	36	−80.129	V	Y	F	H	H	A
12	−83.715	V	V	A	H	A	V	37	−79.958	A	Y	T	H	H	A
13	−83.696	A	Y	F	H	H	V	38	−79.941	A	A	F	H	H	V
14	−83.654	V	V	T	H	T	A	39	−79.925	A	V	T	H	V	V
15	−83.503	A	V	T	H	V	A	40	−79.870	V	Y	F	H	T	A
16	−83.475	A	V	T	H	A	V	41	−79.851	A	Y	T	H	F	A
17	−83.345	A	Y	T	H	T	F	42	−79.838	A	Y	F	H	F	V
18	−83.315	A	Y	T	H	H	F	43	−79.830	A	V	A	H	F	V
19	−83.245	A	V	T	H	T	A	44	−79.816	A	Y	A	H	A	A
20	−83.077	A	Y	F	H	H	F	45	−79.720	A	Y	F	H	H	A
21	−83.033	A	V	T	H	F	F	46	−79.581	A	V	T	H	D	V
22	−82.726	A	V	T	H	H	V	47	−79.500	A	Y	T	H	T	M
23	−82.137	A	V	T	H	T	F	48	−79.461	A	Y	F	H	T	A
24	−82.107	A	V	T	H	H	F	49	−79.346	F	Y	A	H	A	V
25	−81.353	A	V	F	H	H	V	50	−79.322	A	A	F	H	H	F

Table 2. Preferred Residue Selections for Positions 1, 4, 9, 10, 11, and 13 of in Compstatin, as Compared to the Wild-Type Sequence

position	wild type	optimal[a,b]
1	I	A, V
4	V	Y, V
9	H	T, F, A
10	H	H
11	R	T, V, A, F, H
13	T	V, A, F

[a] Only residues with greater than 10 representations among the lowest-lying energy sequences are considered optimal. [b] Listed in decreasing order of optimality.

11 is selected. When considering those positions involved in the hydrophobic cluster of compstatin, it is evident that valine provides strong forces at each position. However, the results for position 4 contrast with those for positions 1 and 13 in that tyrosine, rather than valine, is the preferred choice for the lowest-lying as well as a large majority of the low-lying energy sequences.

It should be noted that, because the compstatin structure was determined via NMR methods, there exists an ensemble of 21 structures for which alternative templates could be derived. These alternative templates were studied as a means of incorporating backbone flexibility into the sequence-selection process, and the results proved to be consistent and in qualitative agreement with those for the average template structure.

3.3. Fold Stability and Specificity Calculations for Selected Sequences. On the basis of the sequence-selection results a handful of optimal sequences were constructed for use in the second stage of the computational design procedure (see Theory and Methods). Because the goal of this work is to gauge the functional improvements of the designed peptides, the first step was to perform the fold stability calculations for the parent peptide sequence (the nonacetylated form of compstatin, for which the NMR structure is available, was used as a basis). For future reference, the fold stability predictions are analyzed according to their relative probabilities using three different classifications. These groupings, denoted as classes, correspond to those sequences exhibiting (i) more than 3 times, (ii) between $^1/_2$ and 3 times, and (iii) less than $^1/_2$ the stability of the parent peptide sequence.

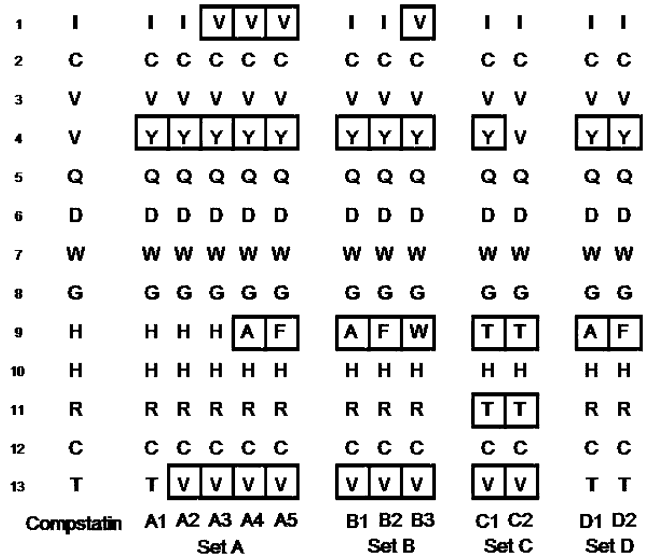

Figure 3. Set of sequences tested for fold stability.

Table 3. Control Sequences Used in Fold Stability Calculations

sequence													
X1	I	C	V	V	Q	D	W	G	A	H	R	C	T
X2	L	C	V	V	Q	D	W	G	W	H	R	C	G

Two additional control experiments were performed to validate these classifications. The sequences used in these trials are shown in Table 3. Sequence X1 corresponds to the conservative substitution of histidine to alanine in position 9, whereas sequence X2 includes a histidine-to-tryptophan substitution in position 9 along with substitutions in both positions 1 and 13. The reasoning for the substitution in sequence X1 is that the introduction of the smaller alanine residues provides additional flexibility for the β-turn,[51] thereby leading to a potential increase in activity. In fact, moderate increases in activity over the wild-type compstatin are observed for both sequence X1 and sequence X2. Fold stability calculations indicate that both sequences belong to class ii, thereby verifying the consistency of the classifications and results.

3.3.1. Significance of Mutations in Set A. For all sequences further characterized via the fold stability calculations (see Figure 3), residue 10 was set to histidine, a prediction consistent with the composition of the parent peptide sequence. Moreover, because the variation in the residue composition for position 11 was predicted to be rather broad, position 11 was restricted to be arginine in subsequent sequences (except set C). The first set of sequences was constructed to better analyze the effect of the tyrosine substitution at position 4, with the justification for focusing on this substitution being an attempt to assess the unusually dominant selection of tyrosine at position 4. The consistent element of the sequences belonging to set A is the assignment of tyrosine to position 4. To further isolate any substitution with respect to the parent peptide sequence, sequences A1−A3 assume the parent compstatin composition of histidine at position 9. Moreover, sequence A1 resembles the parent peptide sequence at positions 1 and 13 as well, whereas sequences A2 and

A3 are constructed so as to add the valine substitutions incrementally, first at position 13 for sequence A2 and then at both positions 1 and 13 for sequence A3. All three sequences fall within class ii according to their fold stability results, with sequences A1 and A3 exhibiting substantial increases in fold stability over the parent peptide sequence. These results highlight the significance of the tyrosine substitution at position 4 and might help to further clarify certain features of the proposed binding model for the compstatin−C3 complex.[51] Specifically, a previous peptide analogue with alanine substitutions on both sides of the β-turn provided activity equivalent to that of the original compstatin sequence. These observations, in addition to the increased turn flexibility afforded by the alanine substitutions, led to the proposition that the presence of side interactions of Val[4] in the binding and activity of compstatin was not required. This corroborates the ability to make the valine-to-tyrosine substitution.

Two additional sequence selections were made to incorporate additional computational and experimental observations. First, on the basis of the results of the control sequence X1, which exhibited increased activity and comparable fold stability when compared to the original compstatin sequence, a similar histidine-to-alanine substitution was made for sequence A4. For sequence A5, position 9 was changed from histidine to phenylalanine, in accordance with the computational results, indicating a strong preference for phenylalanine at this position. On the basis of the fold stability analysis, both sequences were grouped in class ii, with sequence A5 demonstrating the highest relative stability in sequence set A.

3.3.2. Significance of Mutations in Set B. To further explore the combination of position 9 substitutions with the presence of tyrosine at position 4, several additional sequences were constructed. The B1 and B2 constructions represent a reduction in the number of simultaneous mutations from the parent peptide sequence. In effect, these two sequences correspond to the individual combinations of sequence A2 with both sequence A4 and sequence A5 such that position 1 is taken from sequence A2 while position 9 matches the substitutions incorporated into sequences A4 and A5. An additional sequence, B3, is formulated as a combination of sequence A3 and the position 9 substitution of histidine to tryptophan as taken from control sequence X2. Each of the three designed sequences demonstrates a significant increase in fold stability relative to the original compstatin sequence and is classified as belonging to class i.

3.3.3. Significance of Mutations in Set C. Another set of two additional sequences was identified with the only difference between them being the specification of the residue at position 4. For sequence C1, tyrosine was assigned to position 4, whereas sequence C2 was selected to have valine at this position. For both sequences, threonine was specified at positions 9 and 11, and positions 1 and 13 were set to isoleucine and valine, respectively. The choice of isoleucine for position 1 helps to reduce the number of simultaneous changes from the parent peptide sequence.

For both sequence C1 and sequence C2, the stability calculations indicate a substantial decrease in stability when compared to the parent peptide sequence; that is, both sequences belong to class iii. Nevertheless, between sequences C1 and C2, there is strong evidence for the preference of tyrosine at position 4. This prompted closer examination of the residue selections at positions 9 and 11, the two remaining positions not involved in the hydrophobic clustering of compstatin. In particular, the specification of threonine at both positions 9 and 11 results in a negative net charge balance due to the aspartate at position 6, especially because of the replacement of arginine by threonine at position 11. This validates further the placement of arginine at position 11 for the previous set of sequences.

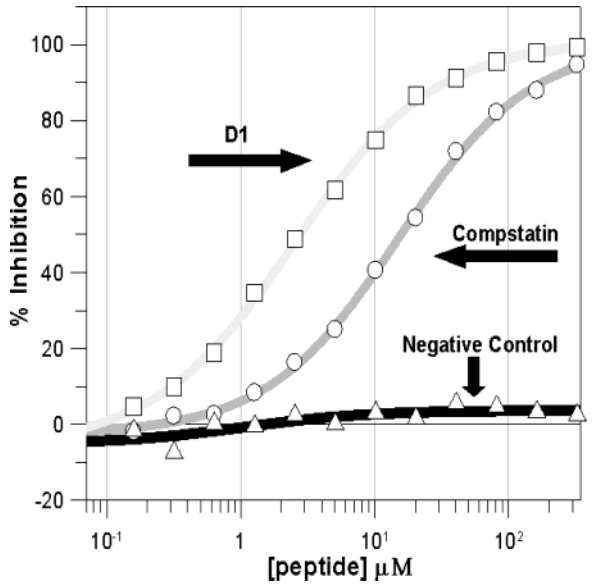

Figure 4. Percent of complement inhibition as a function of peptide concentration for the parent peptide compstatin, the most active analogue D1, and an inactive linear peptide used as a negative control.

3.3.4. Significance of Mutations in Set D. The final set of sequences was designed in accordance with additional reductions in the number of simultaneous mutations relative to the parent peptide sequence. Specifically, sequences D1 and D2 resemble sequences B1 and B2, respectively, with threonine instead of valine as the C-terminal residue, a specification matching the composition of the original parent peptide sequence. Both sequences provide significant increases in fold stability, and are grouped with sequences B1−B3 in class i. For sequences D1 and D2, the differences with respect to the parent peptide sequence are isolated to the single residues before and after the β-turn. Both the position-4 tyrosine and the position-9 phenylalanine substitutions provide enhancements to the fold stability of the compstatin structure and represent unforeseen and unpredictable enhancements over the parent peptide sequence.

3.4. Experimental Studies. A number of the designed sequences presented above were constructed and tested experimentally for their activities, without NMR-based structural analyses being performed. Because the ultimate goal is to enhance the functional activity of compstatin, such achievements must be complemented and verified through experimental studies. Rather than performing massive chemical synthesis of peptide analogues, we tested a few selected analogues against the theoretical predictions. Specifically, one or two analogues were tested experimentally per category in Figure 3. Figure 4 shows the experimentally measured percent complement inhibition as a function of peptide concentration for compstatin, the analogue D1, and the

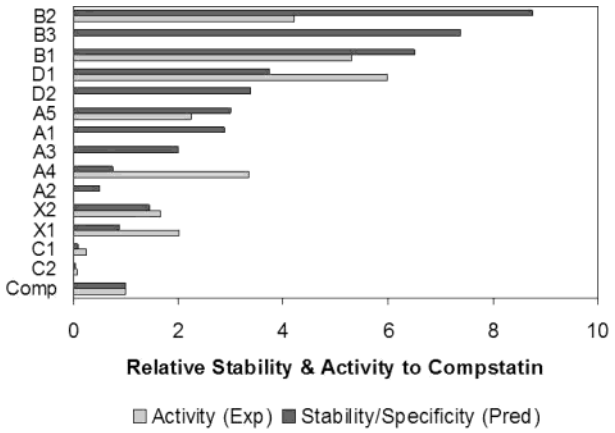

Figure 5. Comparison of relative inhibitory activities and fold stabilities of designed sequences. Theoretical fold stability shown as darkly shaded bars, and experimental inhibitory activity shown as lightly shaded bars. Experimental data are not available for several cases (A1, A2, A3, D2, B3). Three classifications of both relative fold stability and relative activity are made according to the characterization of the parent peptide sequence. The classes represent (i) more than 3 times, (ii) between $^1/_2$ and 3 times, and (iii) less than $^1/_2$ of either the fold stability or the activity of the parent compstatin sequence.

inactive linear analogue C2A/C12A (with Cys2−Ala and Cys12−Ala replacements). Peptide D1 is currently the most active compstatin analogue available. The C2A/C12A analogue is inactive[51] and has been used as a negative control for the inhibition measurements. Figure 5 summarizes the results from the inhibitory activity experiments in comparison to the theoretical fold stability results. Qualitatively, the predicted increases in fold stability and specificity are in excellent agreement with the results from the experimental studies. That is, when considering sets of seuqences (as described above), the computational methodology identifies changes that give rise to real improvements. This is especially significant given that the predictions correspond more directly to fold stability enhancements whereas the experiments directly test inhibitory function.

Although the computational approach is formulated to select and screen sequences in an effort to enhance the fold stability of the template structure, efforts were made to elicit structure−function alterations that would avoid the disruption of elements essential for compstatin activity. The results validate the underlying hypothesis that improved stability and specificity can lead to improved activity when conserving certain functionally important features. This result is significant because such a directed (rational design) formulation can be integrated into an overall predictive method for structure stability and specificity. The comparison between experimental and computational results indicates that the most active compstatin analogues are sequences D1 and B1, as suggested by the optimization study. The common characteristic of these two sequences is the substitutions at positions 4 and 9, the two positions flanking the β-turn residues Gln^5-Asp^6-Trp^7-Gly^8. In particular, the combination of tyrosine at position 4 and alanine at position 9 provides key residues for increased activity and leads to an approximate 6−7-fold improvement over the parent peptide compstatin. This study provides new analogues that outperform analogues identified by either purely rational or experimental combinatorial design studies. In addition, these results are direct evidence that a computational method can be used to predict analogues with enhanced immunological properties.

4. Conclusions

A novel computational structure−activity-based methodology for the de novo design of peptides and proteins was presented. The method is completely general in nature, with the main steps of the approach being obtaining NMR-derived structural templates, combinatorially selecting sequences on the basis of optimization of parametrized pairwise residue interaction potentials, and validating fold stability and specificity using deterministic global optimization. As proof of concept, experimental data were used to formulate a directed design study for the case of the complement inhibitor peptide compstatin. The optimization study led to the identification of many active analogues, including a 6−7-fold more active analogue, as validated through immunological activity measurements. These results are extremely impressive and represent significant enhancements in inhibitory activity over analogues identified by either purely rational or experimental combinatorial design techniques. The work provides direct evidence that an integrated experimental and theoretical approach can make the engineering of compounds with enhanced immunological properties possible.

Acknowledgment

C.A.F. gratefully acknowledges financial support from the National Science Foundation and the National Institutes of Health (R01 GM52032). J.D.L. gratefully acknowledges financial support from the National Institutes of Health (AI 30040 and GM 62134).

Literature Cited

(1) Drexler, K. Molecular engineering: An approach to the development of general capabilities for molecular manipulation. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 5275.

(2) Pabo, C. Moleuclar technology. Designing proteins and peptides. *Nature* **1983**, *301*, 200.

(3) Hardin, C. T. P.; Luthey-Schulten, Z. Ab initio protein structure prediction. *Curr. Opin. Struct. Biol.* **2002**, *12*, 176.

(4) DeGrado, W.; Wasserman, Z.; Lear, J. Protein design, a minimalist approach. *Science* **1989**, *243*, 622.

(5) Hecht, M.; Richardson, D.; Richardson, D.; Ogden, R. De novo design, expression and chracterisation of Felix: A four helix bundle protein with native-like sequence. *Science* **1990**, *249*, 884.

(6) Bowie, J.; Reidhaar-Olson, J.; Lim, W.; Sauer, R. Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* **1990**, *247*, 1306.

(7) Moore, J.; Arnold, F. Directed evolution of a *para*-nitrobenzyl esterase for aqueous−organic solvents. *Nat. Biotechnol.* **1996**, *14*, 458.

(8) Voigt, C.; S. L. Mayo, F. A.; Wang, Z.-G. Computational method to reduce the search space for directed protein evolution. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3778.

(9) Skandalis, A.; Encell, L.; Loeb, L. Creating novel enzymes by applied molecular evolution. *Chem. Biol.* **1997**, *4*, 889.

(10) Ponder, J.; Richards, F. Tertiary templates for proteins. *J. Mol. Biol.* **1987**, *193*, 775.

(11) Hellinga, H.; Richards, F. Construction of new ligand binding sites in proteins of known structure I. Computer aided modeling of sites with predefined geometry. *J. Mol. Biol.* **1991**, *222*, 763.

(12) Desjarlais, J.; Handel, T. De novo design of the hydrophobic cores of proteins. *Protein Sci.* **1995**, *4*, 2006.

(13) Dahiyat, B.; Mayo, S. Protein design automation. *Protein Sci.* **1996**, *5*, 895.

(14) Dahiyat, B.; Gordon, D.; Mayo, S. Automated design of the surface positions of protein helices. *Protein Sci.* **1997**, *6*, 1333.

(15) Raha, K.; Wollacott, A.; Italia, M.; Desjarlais, J. Prediction of amino acid sequence from structure. *Protein Sci.* **2000**, *9*, 1106.

(16) Street, A.; Mayo, S. Pairwise calculation of protein solvent-accessible surface areas. *Fold. Des.* **1998**, *3*, 253.

(17) Nohaile, M.; Hendsch, Z.; Tidor, B.; Sauer, R. Altering dimerization specificity by changes in surface electrostatics. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3109.

(18) Koehl, P.; Levitt, M. De novo protein design I. In search of stability and specificity. *J. Mol. Biol.* **1999**, *293*, 1161.

(19) Wernisch, L.; Hery, S.; Wodak, S. Automatic protein design with all atom force-fields by exact and heuristic optimization. *J. Mol. Biol.* **2000**, *301*, 713.

(20) Jones, D. De novo protein design using pairwise potentials and a genetic algorithm. *Protein Sci.* **1994**, *3*, 567.

(21) Desjarlais, J.; Handel, T. Side chain and backbone flexibility in protein core design. *J. Mol. Biol.* **1999**, *290*, 305.

(22) Voigt, C.; Gordon, D.; Mayo, S. Trading accuracy for speed: a quantitative comparison of search algorithms in protein sequence design. *J. Mol. Biol.* **2000**, *299*, 789.

(23) Desmet, J.; Maeyer, M. D.; Hazes, B.; Lasters, I. The dead-end elimination theorem and its use in side-chain positioning. *Nature* **1992**, *356*, 539.

(24) Gordon, D.; Mayo, S. Branch-and-terminate. *Struct. Fold. Des.* **1999**, *7*, 1089.

(25) Jin, W.; Kambara, O.; Sasakawa, H.; Tamura, A.; Takada, S. *Structure* **2003**, *11*, 581.

(26) Dahiyat, B.; Mayo, S. De novo protein design: Fully automated sequence selection. *Science* **1997**, *278*, 82.

(27) Klepeis, J. E.; Floudas, C. A.; Morikis, D.; Tsokos, C. G.; Argyropoulos, E.; Spruce, L.; Lambris, J. D. Integrated Computational and Experimental Approach for Lead Optimization and Design of Compstatin Variants with Improved Activity. *J. Am. Chem. Soc.* **2003**, *125*, 8422.

(28) Morikis, D.; Assa-Munt, N.; Sahu, A.; Lambris, J. D. Solution Structure of Compstatin, a Potent Complement Inhibitor. *Protein Sci.* **1998**, *7*, 619.

(29) Park, B.; Levitt, M. Energy functions that discriminate X-ray and near native folds from well-constructed decoys. *J. Mol. Biol.* **1996**, *258*, 367.

(30) Meller, J.; Elber, R. Linear programming optimization and a double statistical filter for protein threading protocols. *Proteins* **2001**, *45*, 241.

(31) Tobi, D.; Elber, R. Distance-dependent pair potential for protein folding: Results from linear optimization. *Proteins* **2000**, *41*, 40.

(32) Tobi, D.; Shafran, G.; Linial, N.; Elber, R. On the design and analysis of protein folding potentials. *Proteins* **2000**, *40*, 71.

(33) Loose, C.; Klepeis, J.; Floudas, C. A new pairwise folding potential based on improved decoy generation and side chain packing *Proteins* **2004**, *54*, 303.

(34) Floudas, C. A. *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*; Oxford University Press: New York, 1995.

(35) *Using the CPLEX Callable Library*; ILOG, Inc.: Mountain View, CA, 1997.

(36) Sherali, H.; Adams, W. *A Reformulation Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*; Kluwer Academic Publishers: Boston, 1999.

(37) Klepeis, J. L.; Schafroth, H. D.; Westerberg, K. M.; Floudas, C. A. Deterministic Global Optimization and ab Initio Approaches for the Structure Prediction of Polypeptides, Dynamics of Protein Folding and Protein−Protein Interaction. In *Advances in Chemical Physics*; Friesner, R. A., Ed.; Wiley: New York, 2002; Vol. 120, pp 254−457.

(38) Klepeis, J. L.; Floudas, C. A.; Morikis, D.; Lambris, J. Predicting Peptide Structures Using NMR Data and Deterministic Global Optimization. *J. Comput. Chem.* **1999**, *20*, 1354.

(39) Klepeis, J.; Floudas, C. Ab initio tertiary structure prediction of proteins. *J. Global Optim.* **2003**, *25*, 113.

(40) Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. Energy Parameters in Polypeptides. 10. *J. Phys. Chem.* **1992**, *96*, 6472.

(41) Adjiman, C.; Androulakis, I.; Floudas, C. A. A Global Optimization Method, αBB, for General Twice-Differential Constrained NPLs − I. Theoretical Advances. *Comput. Chem. Eng.* **1998**, *22*, 1137.

(42) Adjiman, C.; Androulakis, I.; Floudas, C. A. A Global Optimization Method, αBB, for General Twice-Differentiable Constrained NLPs − II. Implementation and Computational Results. *Comput. Chem. Eng.* **1998**, *22*, 1159.

(43) Adjiman, C.; Androulakis, I.; Floudas, C. A. Global Optimization of Mixed-Integer Nonlinear Problems. *AIChE J.* **2000**, *46*, 1769.

(44) Klepeis, J. L.; Floudas, C. A. Free Energy Calculations for Peptides via Deterministic Global Optimization. *J. Chem. Phys.* **1999**, *110*, 7491.

(45) Floudas, C. A. *Deterministic Global Optimization: Theory, Methods and Applications: Nonconvex Optimization and Its Applications*; Kluwer Academic Publishers: Boston, 2000.

(46) Klepeis, J.; Pieja, M.; Floudas, C. A New Class of Hybrid Global Optimization Algorithms for Peptide Structure Prediction: Integrated Hybrids. *Comput. Phys. Commun.* **2003**, *151*, 121.

(47) Klepeis, J.; Pieja, M.; Floudas, C. Hybrid Global Optimization Algorithms for Protein Structure Prediction: Alternating Hybrids. *Biophys. J.* **2003**, *84*, 869.

(48) Sahu, A.; Kay, B.; Lambris, J. Inhibition of human complement by a C3-binding peptide isolated from a phage displayed random peptide library. *J. Immunol.* **1996**, *157*, 884.

(49) Sahu, A.; Soulika, A.; Morikis, D.; Spruce, L.; Moore, W.; Lambris, J. Binding kinetics, structure activity relationship and biotransformation of the complement inhibitor compstatin. *J. Immunol.* **2000**, *165*, 2491.

(50) Sahu, A.; Lambris, J. Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. *Immunol. Rev.* **2001**, *180*, 35.

(51) Morikis, D.; Roy, M.; Sahu, A.; Torganis, A.; Jennings, P.; Tsokos, G.; Lambris, J. The structural basis of compstatin activity examined by structure−function-based design of peptide analogues and NMR. *J. Biol. Chem.* **2002**, *277*, 14942.

Received for review August 29, 2003
Revised manuscript received January 25, 2004
Accepted January 28, 2004

IE0340995