Structure-Based Integrative Computational and Experimental Approach for the Optimization of Drug Design*

Dimitrios Morikis[1], Christodoulos A. Floudas[2], and John D. Lambris[3]

[1] Department of Chemical and Environmental Engineering, University of California, Riverside, CA 92521, USA
dmorikis@engr.ucr.edu
[2] Department of Chemical Engineering, Princeton University, Princeton, NJ 08544, USA
floudas@titan.princeton.edu
[3] Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
lambris@mail.med.upenn.edu

Abstract. We present an integrative approach for the optimization in the design of peptides which are candidates to become therapeutic agents. This approach is based on the structure of the peptide ligand when structural information on the protein target is not available. Our approach combines (i) NMR spectroscopy, (ii) structure determination by distance geometry, simulated annealing, and global optimization methods, restrained with NMR-derived or deduced restraints, (iii) molecular dynamics simulations, based on NMR low energy, averaged minimized, or ensemble of structures, (iv) in silico sequence selection using integer linear optimization, (v) fold specificity using deterministic global optimization, and (vi) peptide synthesis, mass spectrometry characterization, and activity measurements. The optimization of the design of the 13-residue cyclic peptide compstatin is presented as a paradigm for the application of our approach. The same principles can be applied for the design of small proteins with desired properties and function.

1 Introduction

We present an overview of our integrative approach for peptide-drug discovery. This approach is based on the peptide structure and it incorporates classical and novel features for drug design methodologies. The approach integrates several methods, computational and experimental, for structural analysis, structure determination, sequence and structure prediction, and structure-activity correlations, including spectroscopy, spectrometry, molecular dynamics simulations, integer linear optimization and global optimization, and binding constant and IC_{50}-value activity

* This work was supported by grants from NIH and NSF.

V.S. Sunderam et al. (Eds.): ICCS 2005, LNCS 3515, pp. 680 – 688, 2005.

measurements. We will present brief descriptions for the use and integration of the computational and experimental methods. We will follow with a specific example for the application of our approach on the optimization of the peptide compstatin, a potent inhibitor against the abnormal (or unregulated) function of the complement system (part of innate immunity).

Our approach can be used for the optimization in the design of active peptide analogs, with the aim to improve their activity. Optimally, the structures of both the ligand peptide and the target protein or the active (binding) site of the target protein are desirable. In this paper we focus on ligand-based design, using the structure of the peptide when the structure of the protein is not available. This is often the case for large or unstable proteins, which are not yet amenable to structure determination by crystallography (X-ray or neutron diffraction) or NMR methods.

2 Methods

Figure 1 presents the interplay of the computational and experimental methods used in our approach. The methods used in the various steps will be briefly described below.

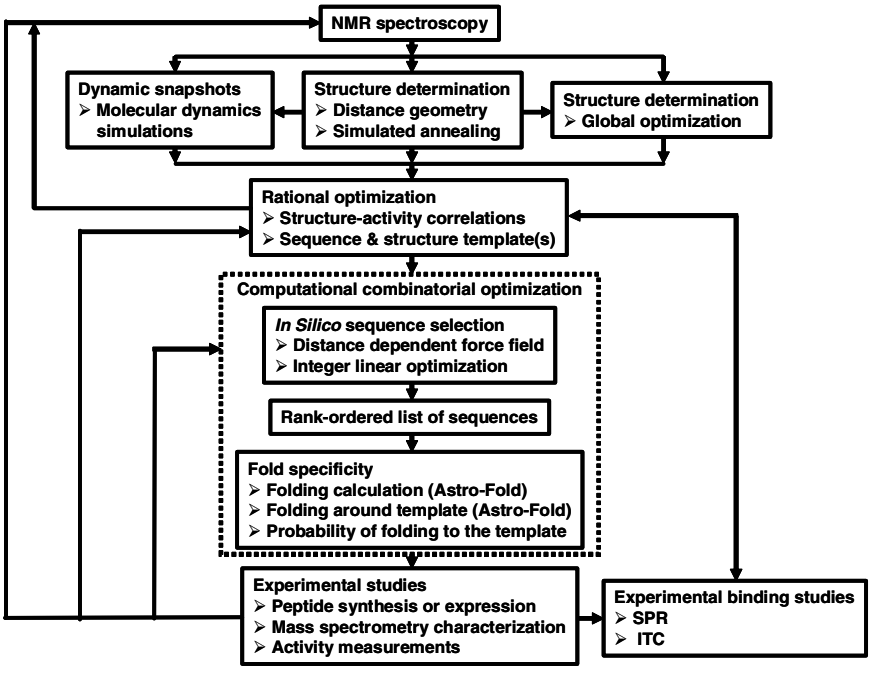

Fig. 1. Flow of information in our integrative drug design and optimization approach

2.1 NMR Spectroscopy

Multi-dimensional and multi-nuclear NMR spectroscopy is used to determine the secondary structure and tertiary or quaternary contacts of peptides and proteins [1]. Typically, two-dimensional NMR spectra are sufficient for structural analysis of peptides. In cases of overlapping cross peaks three-dimensional heteronuclear NMR spectra may be necessary to increase spectral resolution. Heteronuclear NMR spectra require, in most cases, ^{15}N- and or ^{13}C-labeled samples, prepared by expression rather than chemical synthesis methods.

The measured NMR parameters are chemical shifts, spectral linewidths and heights, areas, or volumes, coupling constants, and nuclear Overhauser effects (NOEs). The chemical shifts are used to assign specific resonances for protons and ^{13}C and ^{15}N. Patterns of chemical shifts are used to identify the amino acid systems. Deviations of chemical shifts in structured peptides from their random coil values, plotted as differences or chemical shift indices, are used to distinguish helical, beta-strand, and random coil secondary structures. Temperature variation of chemical shifts often points to the formation of hydrogen bonded secondary structure, when plotted in the form of temperature coefficients. In certain triple resonance NMR spectra, connectivities of chemical shifts are used to identify the protein backbone by piecing together the backbone N, C_α, C' atoms, and side chain C_β atoms. Coupling constants are also used to distinguish well-formed helical from beta-strand secondary structures. NOEs are used to determine through space dipolar interactions of protons in proximity of less than 5.5-6 Å. Specific sequential and medium range NOE connectivity patterns are used to determine secondary structure and to distinguish helices, strands, and turns, from extended or random conformations. Long range NOEs are used to identify tertiary and inter-molecular contacts in the case of complexes. Spectral linewidths are used to determine relaxation parameters and to identify aggregation or binding.

2.2 Structure Determination Using NMR Restraints

Distances and backbone and side chain torsion angles are typically used as experimental restraints, together with restraints from chemical knowledge for covalent geometry and nonbonded interactions, in computational methods for the determination of three-dimensional structures of peptides [2]. Distances are derived from NOEs and torsion angles are derived from J-coupling constants, sometimes in combination with NOEs. The structure determination protocols are based on the minimization of an energy function using parameters and topologies from a specific force field, in the Cartesian or torsion angle space.

Distance Geometry and Molecular Dynamics-Based Simulated Annealing. Distance geometry methods are based on the metric matrix theorem which allows for the conversion of distances among a set of points in three-dimensional Euclidian space into Cartesian coordinates. Because not all atom pair distances are

known in the metric matrix, large distances need to be assigned and subsequently reduced using the triangle inequality. In some instances, the distance geometry method is the first step in structure determination followed by simulated annealing regularization of the structures. Alternatively, simulated annealing methods are used alone.

Molecular dynamics-based simulated annealing for the determination of three-dimensional structures using NMR restraints is a simplified implementation of regular (unrestrained) molecular dynamics *in vacuum* (without the presence of explicit solvent molecules or implicit solvation energy term) [2]. The potential energy function has the form $E=E_{covalent}+E_{nonbonded}+E_{experimental}$, where $E_{covalent}=E_{bond}+E_{angle}+E_{torsion}+E_{chiral,planar}$, and $E_{experimental}=E_{distance}+E_{torsion}+(E_{chemicalshift}+E_{couplingconstant})$. The terms in parentheses are optional or entered during the later stages of the calculations. $E_{covalent}$ and $E_{nonbonded}$ are empirical energy terms from known chemical principles. $E_{nonbonded}$ is typically a simplified repel potential term that accounts for van der Waals and electrostatic interactions of nonbonded atom pairs. In certain instances or at later stages in the calculations, deduced hydrogen bonds can be used in the form of distance restraints. The structure determination protocols are iterative procedures involving the stepwise addition of restraints and tests for the correctness of the assigned restraints.

Deterministic Global Optimization. Global optimization methods have been tested first for the structure determination of peptides using NOE restraints [3]. The structure determination formulation represents a general nonconvex constrained global optimization problem, a class of problems for which several 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 [4]. The global minimization of a detailed atomistic energy force field $E_{forcefield}$ is performed over the set of independent torsion angles, which can be used to describe any possible configuration of the system. The bounds on the torsion angles are enforced by simple box restraints. Finally, a set of distance constraints, which are nonconvex in the internal coordinate system, are used to constrain the system. The energy function has the form $E_{forcefield}=E_{electrostatic}+E_{vanderWaals}+E_{Hbond}+E_{torsion}$.

2.3 Molecular Dynamics Simulations

Molecular dynamics simulations using implicit solvent representation are performed to scan rapidly the conformational space of peptides and proteins [1]. The initial structures are the complete ensemble of NMR-derived structures, the averaged minimized structure, or the lowest energy structure. The empirical energy function has the form $E=E_{covalent}+E_{nonbonded}+E_{solvation}$, where $E_{covalent}$ includes the covalent geometry energy terms $E_{bond}+E_{angle}+E_{torsion}+E_{chiral,planar}$, $E_{nonbonded}$ includes the nonbonded interaction terms $E_{vanderWaals}+E_{electorstatic}+(E_{Hbond})$, and $E_{solvation}$ is an appropriate impicit solvation model energy term. Simulation times in the range of 1-

10 ns are typically sufficient for peptides. Energy minimizations are performed before and after the molecular dynamics simulations.

Individual snapshots during a trajectory are analyzed to examine (i) flexibility using the RMSD of backbone and side chain atoms and calculated B-factors, (ii) secondary structure using backbone torsion angles and hydrogen bonding, and (iii) tertiary structure using side chain contacts. Pairwise nonbonded interactions between side chains are determined by plotting the total potential energy, the van der Waals energy, and electrostatic energy. This type of analysis evaluates the role of side chains in packing and stability. Molecular dynamics also provide free energy differences of different conformations and motional amplitudes for conformational inter-conversion.

2.4 Rational Design and Optimization

Rational design and optimization depends on the availability of three-dimensional structures [5,6]. In rational design the structures or structural information of active analogs from NMR and computational methods (see above) are used to identify the critical for binding and activity physico-chemical properties and their spatial arrangement. This is done in combination with experimental binding and activity measurements (see below) by determining structure-(dynamics-binding)-activity correlations. The parentheses denote often optional steps; however these steps are necessary in our opinion because in several cases the lowest energy structure derived from NMR or crystallography is not the one that binds. It is not unusual for the ligand or the protein active site to undergo structural rearrangements upon binding. Structure-(dynamics-binding)-activity correlations are used to optimize the peptide by rationally replacing or modifying its building blocks. Ligand-based Pharmacophore and QSAR (quantitative structure-activity relationship) approaches use rational design for the construction of pharmacophore models. Peptide pharmacophore models can be used to identify matching low molecular mass organic molecules from appropriate databases. It is usual in rational design to incorporate non-natural amino acids or amino acids substituted with specific chemical groups (e.g., methylation, etc), which need parameterization for the computational methods described below.

2.5 Computational Combinatorial Design and Optimization

A two-stage method for the de novo peptide and protein design has been recently introduced. This method is based on sequence and structural templates of active analogs, determined from NMR- or crystallographically-derived structures, or from snapshots of molecular dynamics trajectories. The first step of the method involves sequence selection and the second step involves fold validation.

Sequence Selection. The sequence selection step relies on a constrained integer linear programming (ILP) model [7,8]. Sequences compatible with given sequence and backbone templates are identified by minimizing an empirical potential describing

pairwise distance-dependent interactions. This potential assigns energy values for amino acid interactions, based on the C_α-C_α separation distance for each amino acid pair. The type of amino acids (side chains) is implicitly included in the interaction potentials. The solution of this ILP problem allows the identification of a rank ordered list of the low lying energy sequences, which are used in the second step, the prediction of fold stability and specificity, described next.

Prediction of Fold Specificity. This step is used to distinguish the most optimal sequences from those identified in the sequence selection step (above) according to rigorous quantification of conformational probabilities [7,8]. Conformational ensembles are generated for the selected sequences under two sets of conditions. First, the structure is allowed to vary around the template structure, with imposed fluctuations. The fluctuations can be based on the structural boundaries defined by the NMR ensemble, on the RMSDs of molecular dynamics structures, or some arbitrary fluctuation. Second, a free folding calculation is performed with limited number of restraints, as needed (e.g., disulfide bridges, etc), but with the underlying template structure not being enforced. The ensembles of conformers resulting from the two folding calculations are used to calculate the probabilities. The formulation of the folding calculations is similar to the structure determination calculations using deterministic global optimization (described above). The calculations are performed with ASTRO-FOLD framework [9].

2.6 Experimental Peptide Synthesis and Binding and Activity Measurements

Sample preparation is typically performed using solid state peptide synthesis or expression methods and tested for integrity using mass spectrometry. Biological assays for activity measurement in the form of IC_{50} values are used to assess the strength of the peptide inhibitors. The IC_{50} value is the peptide concentration at 50% inhibition. Direct inhibition or competition assays using ELISA methods are used, depending on the specifics of the experiment. Binding data using surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) methods are often necessary to correlate structure with binding and activity and to form hypotheses [1].

3 Results: The Example of Compstatin

The 13-residue peptide compstatin is an inhibitor of the complement system, with sequence I[CVVQDWGHHRC]T-NH_2, where brackets denote cyclization through a disulfide bridge [5]. The complement system is part of innate immunity and acts together with adaptive immunity to fight infection by foreign pathogens. The complement system is finely regulated by fluid phase or cell-bound regulators and is programmed to recognize “self” from “non-self”. When this regulation brakes down,

as is the case in several autoimmune diseases and pathological situations, the complement system turns against "self" and attacks host tissues. Currently there is no drug in the clinic that regulates the abnormal complement activation.

Compstatin was identified by constructing and testing a phage-displayed random peptide library against binding to complement component C3 [10], a converging component of the complex pathways of complement activation. The three-dimensional structure of compstatin was determined using NMR data and computational methods restrained with NMR-derived restraints. The computational methods were hybrid distance geometry/simulated annealing [11] and global optimization [3]. Subsequently, optimization of the sequence of compstatin was performed using rational design based on NMR structural studies (but not computational complete structure determination) and structure-activity correlations, which yielded several active analogs with up to 4-fold higher inhibitory activity than the parent peptide [6]. The rational design determined that 7 of the 13 amino acids of compstatin were indispensable for activity, and provided the following sequence template for further optimization: Ac-**X**[CV**X**QDWG**XXX**C]**X**-NH_2 (called active sequence template), where the 6 amino acids marked with **X** were optimizable [5]. The active sequence template was used for the construction of a second round of phage-displayed peptide library and binding experiments against C3 (called experimental combinatorial optimization). This round of optimization yielded four more active analogs, with one of them being 4-fold more active than the parent peptide [12]. The active sequence template and the NMR-derived structure of compstatin were also used as the first test case of the novel computational combinatorial approach described above, which predicted several active analogs [7,8]. Among the active analogs that were synthesized and experimentally tested for activity were 5 analogs with 6- to 14-fold higher inhibitory activities than the parent peptide [7,8,13]. Subsequent rounds of rational design, using the same active sequence template and NMR-based structural studies identified several active analogs with up to 99-fold higher inhibitory activities than the parent peptide [13]. Several of these analogs are peptidomimetics because they are built using combinations of natural and non-natural amino acids. Table 1 shows the major breakthroughs in the optimization of the design of compstatin, from each of the various optimization rounds.

Besides IC_{50} activity measurements that were performed for each synthesized analog, kinetic and thermodynamic binding studies were performed using surface plasmon resonance [12,14], and isothermal titration calorimetry [15]. The structural, binding, and activity studies were useful to form testable structure-binding and structure-activity hypotheses.

Finally, quasi-dynamic pharmacophore models have been generated using snapshots from molecular dynamics simulations of several active and inactive compstatin analogs [16,17]. Upon selection of proper geometric and physico-chemical properties to represent the spatial arrangement of the pharmacophore points, one model allowed for the distinction of active from inactive analogs [17].

Table 1. Benchmarks in the optimization of the design of compstatin*

Peptide	Sequence	RIA	Year	Ref.
I	I[CV V QDWGHHRC]T-NH_2	1	1996	10
II	**Ac**-I[CV V QDWG**A**HRC]T-NH_2	4	2002	6
III	**Ac**-**L**[CV V QDWG**W**HRC]**G**-NH_2	4	2003	12
IV	**Ac**-I[CV **Y** QDWG**A**HRC]T-NH_2	14	2003	7,8
V	**Ac**-I[CV **W** QDWG**A**HRC]T-NH_2	45	2005	13
VI	**Ac**-I[CV(**2Nal**)QDWG**A**HRC]T-NH_2	99	2005	13

*RIA, relative inhibitory activity. Bold face indicates additions/substitutions responsible for the increase in inhibitory activity. Ac, acetylation; 2Nal, 2-naphthylalanine.

4 Conclusions

We have described a ligand-based approach for the optimization of active peptides, which are candidates to become therapeutics. This approach involves the interplay of a variety of computational and experimental methods. The compstatin example presented here, has demonstrated the effectiveness of our ligand-based approach in the absence of structural information for the C3 target, the structure of which thus far has not been determined. We expect that similar mixed computational and experimental approaches will also be useful when the structures of both, the free ligand and the free target, and/or their complex, are available.

References

1. Morikis, D., and Lambris, J.D.: Physical methods for structure, dynamics and binding in immunological research. Trends Immunol. 25 (2004) 700-707
2. Güntert, P.: Structure calculation of biological macromolecules from NMR data. Quart. Rev. Bioph. 31 (1998) 145-237
3. Klepeis, J.L., Floudas, C.A., Morikis, D., Lambris, J.D.: Predicting peptide structures using NMR data and deterministic global optimization. J. Comp. Chem. 20 (1999) 1354-1370
4. Floudas, C.A.: Deterministic global optimization: theory, methods and applications, Kluwer Academic Publishers, Dordrecht, The Nederlands (2000)
5. Morikis, D., Soulika, A.M., Mallik, B., Klepeis, J.L., Floudas, C.A., and Lambris, J.D.: Improvement of the anti-C3 activity of compstatin using rational and combinatorial approaches. Biochem. Soc. Trans. 32 (2004) 28-32
6. Morikis, D., Roy, M., Sahu, A., Troganis, A., Jennings, P.A., Tsokos, G.C., and Lambris J.D.: The structural basis of compstatin activity examined by structure-function-based design of peptide analogs and NMR. J. Biol. Chem. 277 (2002) 14942-14953
7. Klepeis, J.L., Floudas, C.A., Morikis, D., Tsokos, C.G., Argyropoulos, E., Spruce, L.A., Lambris, J.D.: Integrated computational and experimental approach for lead optimization and design of compstatin variants with improved activity. J. Am. Chem. Soc. 125 (2003) 8422-8423
8. Klepeis, J.L., Floudas, C.A., Morikis, D., Tsokos, C.G., and Lambris, J.D.: Design of peptide analogues with improved activity using a novel de novo protein design approach. Ind. Eng. Chem. Res. 43 (2004) 3817-3826

9. Klepeis, J.L. and Floudas, C.A.: ASTRO-FOLD: a combinatorial and global optimization framework for ab initio prediction of three-dimensional structures of proteins from the amino acid sequence. Biophys. J. 85 (2003) 2119-2146
10. Sahu, A., Kay, B.K., Lambris, J.D.: Inhibition of human complement by a C3-binding peptide isolated from a phage-displayed random peptide library. J. Immunol. 157 (1996) 884-891
11. Morikis, D., Assa-Munt, N., Sahu, A., Lambris, J.D.: Solution structure of compstatin, a potent complement inhibitor. Protein Sci. 7 (1998) 619-627
12. Soulika, A.M., Morikis, D., Sarrias, M.R., Roy, M., Spruce, L.A., Sahu, A., Lambris, J.D.: Studies of Structure-Activity Relations of Complement Inhibitor Compstatin. J. Immunol. 171 (2003) 1881-1890
13. Mallik, B., Katragadda, M., Spruce, L.A., Carafides, C., Tsokos, C.G., Morikis, D., and Lambris J.D.: Design and NMR characterization of active analogs of compstatin containing non-natural amino acids. J. Med. Chem. 48 (2005) 274-286
14. Sahu, A., Soulika, A.M., Morikis, D., Spruce, L.A., Moore, W.T., and Lambris, J.D.: Binding kinetics, structure-activity relationship, and biotransformation of the complement inhibitor Compstatin. J. Immunol. 165 (2000) 2491-2499
15. Katragadda, M., Morikis, D., and Lambris, J.D.: Thermodynamics studies on the interaction of the third complement component and its inhibitor, compstatin. J. Biol. Chem. 279 (2005) 54987-54995
16. Mallik, B., Lambris, J.D., Morikis, D.: Conformational inter-conversion of compstatin probed with molecular dynamics simulations. Proteins 53 (2003) 130-141
17. Mallik, B. and Morikis D.: Submitted (2005)