1-1-2012

Computational Protein Design and Molecular Dynamics Simulations: A Study of Membrane Proteins, Small Peptides and Molecular Systems

Jose Manuel Perez Aguilar
University of Pennsylvania, perezagu@sas.upenn.edu

Recommended Citation

Perez Aguilar, Jose Manuel, "Computational Protein Design and Molecular Dynamics Simulations: A Study of Membrane Proteins, Small Peptides and Molecular Systems" (2012). *Publicly Accessible Penn Dissertations*. 480.
http://repository.upenn.edu/edissertations/480

Computational Protein Design and Molecular Dynamics Simulations: A Study of Membrane Proteins, Small Peptides and Molecular Systems

Abstract

Molecular design and modeling can provide stringent assessment of our understanding of the structure and function of proteins. Due to the subtleness of the interactions that largely stabilize proteins, computational methods have been particularly valuable in establishing practical, formal and physically grounded protocols to study the structure and function of these biomolecules. Especifically, computational protein design seeks to identify sequences that fold into a desired structure and have specific structural and functional properties using computational methodologies. Among current techniques, an entropy-based formalism that efficiently determines the number and composition of sequences satisfying a predefined set of constraints seems particularly promising and powerful. Complementary to this methodology are the well-established molecular dynamics simulation techniques that have been extensively used to study structure, function and dynamics of biologically relevant systems. Herein different studies of systems using computational techniques to address particular molecular problems are described. Efforts to redesign membrane proteins to generate water-soluble variants were applied to a widely studied pentameric ligand-gated ion channel, the nicotinic acetylchoilne receptor (nAChR). NMR structures and binding studies demostrated the robustness and applicability of the computational design approach. Toward the creation of water-soluble variants of a G protein–coupled receptor (GPCR), comparative modeling and docking calculations were used to investigate the structure of the human μ opioid receptor and presented in light of previous mutagenesis studies of structure and agonist-induced activation. Candidate peptides for possible therapeutic agents were computationally analyzed. Peptide design, loop modeling and MD simulations were applied to investigate the stromal cell-derived factor-1&a; (SDF-1&a;). SDF-1&a; displays promising therapeutic benefits to treat blood-supply related heart disease and elicit growth of microvasculature. Simplified analogs of SDF-1&a; exhibit enhanced therapeutic properties in cell-based assays. MD simulations provide insights about the molecular features of this enhancement. One simplified peptide offers a potentially clinically translatable neovasculogenic therapy. Lastly, MD simulations were utilized to analyze a molecule with hindered internal rotors, a tribenzylamine hemicryptophane. The molecule was characterized by different experimental and computational techniques. The structural and dynamic features of the hemicryptophane molecule make it an attractive starting point for controlling internal rotation of aromatic rings within molecular systems.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Chemistry

First Advisor
Jeffery G. Saven

Keywords
Computational Protein Design, Entropy-Based Formalism, Molecular Dynamics Simulation, Statistical Mechanics

Subject Categories
Biophysics | Chemistry

COMPUTATIONAL PROTEIN DESIGN AND MOLECULAR DYNAMICS SIMULATIONS: A STUDY OF MEMBRANE PROTEINS, SMALL PEPTIDES AND MOLECULAR SYSTEMS

JOSÉ MANUEL PÉREZ AGUILAR

A DISSERTATION
in
Chemistry

Presented to the Faculties of the University of Pennsylvania
in
Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
2012

Supervisor of Dissertation

Jeffery G. Saven, Associate Professor of Chemistry

Graduate Group Chairperson

Gary A. Molander, Professor of Chemistry

Dissertation Committee

Feng Gai, Professor of Chemistry
Michael L. Klein, Professor of Chemistry, Temple University
Tobias Baumgart, Associate Professor of Chemistry

DEDICATION

to Brenda

“Nature never undertakes any change unless her interests are served by an increase in entropy.”

– Max Planck, 1903

ACKNOWLEDGEMENTS

I like to express my deepest gratitude to my advisor, Professor Jeffery G. Saven, for taking me into his research group. His guidance and support nurtured me as a scientist through my graduate life. His mentorship showed me scientific and moral principles that are invaluable beyond my professional career. I feel grateful and proud for sharing all these years under his supervision, which have prepared me to pursuit all my scientific endeavors.

To my committee members, Professor Feng Gai, Professor Tobias Baumgart and Professor Michael L. Klein, thank you for you advice and direction during all these years. Thank you for your supervision and fruitful discussions that challenged my knowledge and enhanced the understanding of my scientific projects.

My profound appreciation to all my collaborators, thank you for your hard work and for sharing with me some of your knowledge. Professor Roderick G. Eckhenhoff and his group at the University of Pennsylvania. Professor Yan Xu and his group at the University of Pittsburgh. Professor Renyu Liu and his group at the University of Pennsylvania. Professor Y. Joseph Woo and MD. Willliam Hiesinger at the University of Pennsylvania. Professor Ivan J. Dmochowski and Dr. Najat S. Khan at the University of Pennsylvania. Professor Michael J. Therien and his group at Duke University. Professor Hiroaki Matsunami and his group at Duke University.

I would like to thank to all the current and former members of the Saven research group for sharing many things during my time in the group and especially for their friendship. Dr. Christopher L. Lanci, Christopher M. MacDermaid, Christopher D. Von Bargen, Lu Gao and Matthew J. Eibling. Also to Dr. One-sun Lee, Dr. Andreas Lehmann, Dr. Ilan Samish, Dr. Thomas J. Petty, Dr. Feng Chen, Dr. Seung-gu Kang, MS. Lin Zhao and Anand Gopal.

I wish to express my love and gratitude to my mother, Virginia Aguilar Fernandez, my grandparents Manuela y Alfonso, my aunts Laura and Tere, and my entire family for their encouragement, support, and endless love during all my professional and personal endeavors. Finally, I want to thank to the love of my life Brenda Leonor Sanchez Gaytan, for her friendship, support, and existence, and for deciding to walk with me this journey called life.

ABSTRACT

COMPUTATIONAL PROTEIN DESIGN AND MOLECULAR DYNAMICS SIMULATIONS: A STUDY OF MEMBRANE PROTEINS, SMALL PEPTIDES AND MOLECULAR SYSTEMS

José Manuel Pérez Aguilar

Jeffery G. Saven

Molecular design and modeling can provide stringent assessment of our understanding of the structure and function of proteins. Due to the subtleness of the interactions that largely stabilize proteins, computational methods have been particularly valuable in establishing practical, formal and physically grounded protocols to study the structure and function of these biomolecules. Especifically, computational protein design seeks to identify sequences that fold into a desired structure and have specific structural and functional properties using computational methodologies. Among current techniques, an entropy-based formalism that efficiently determines the number and composition of sequences satisfying a predefined set of constraints seems particularly promising and powerful. Complementary to this methodology are the well-established molecular dynamics simulation techniques that have been extensively used to study structure, function and dynamics of biologically relevant systems. Herein different studies of systems using computational techniques to address particular molecular problems are described. Efforts to redesign membrane proteins to generate water-soluble variants were applied to a widely studied pentameric ligand-gated ion channel, the nicotinic acetylchoilne receptor (nAChR). NMR structures and binding studies demostrated the robustness and applicability of the computational design approach. Toward the creation of water-soluble variants of a G protein−coupled receptor (GPCR), comparative modeling and docking calculations were used to investigate the structure of the human μ opioid receptor and presented in light of previous mutagenesis studies of structure and agonist-induced activation. Candidate peptides for possible therapeutic agents were computationally analyzed. Peptide design, loop modeling and MD simulations were applied to investigate the stromal cell–derived factor–1α (SDF–1α). SDF–1α displays promising therapeutic benefits to treat blood-supply related heart disease and elicit growth of microvasculature. Simplified analogs of SDF–1α exhibit enhanced therapeutic properties in cell-based assays. MD simulations provide insights about the molecular features of this enhancement. One simplified peptide offers a potentially clinically translatable neovasculogenic therapy. Lastly, MD simulations were utilized to analyze a molecule with hindered internal rotors, a tribenzylamine hemicryptophane. The molecule was characterized by different experimental and computational techniques. The structural and dynamic features of the hemicryptophane molecule make it an attractive starting point for controlling internal rotation of aromatic rings within molecular systems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

Proteins play crucial roles in the cellular physiology of a living organism with a wide diversity of biological functions, including storage and transport of different particles, catalysis (enzymes), transmission of information (hormones), control of flux of molecules across the membrane, immune responses, control of gene expression, and structural support [1].

To have biological activity, proteins must adopt specific folded three-dimensional conformations, also denominated native conformations. The three-dimensional structure is determined by the linear sequence of amino acids in the protein backbone [2]. The description of the spontaneous formation of a unique highly ordered three-dimensional structure from the complete unfolded polypeptide chain comformations has been denominated the protein folding problem [3].

Complementary to the protein folding problem is the approach of finding an “appropriate” set of amino acid sequences that fold into a desired target structure in a similar way as a native conformation with the lowest accessible free energy. Furthermore, such “appropriate” sequences will not simultaneously fold to alternative competing conformations of similar free energy [4,5]. In other words, instead of starting with a specific amino acid sequence and then predict the folded structure of the native conformations, the protein design paradigm starts with a conformation of the backbone and then selects amino acid sequences that stabilize the native three-dimensional conformation. This inverse approach, sometimes denominated inverse protein folding problem [4,5], is the purpose of the current efforts of the protein design field [6-9].

1.1 Computational Protein Design

Early protein design efforts involved the design of small proteins using empirical knowledge obtained from structural databases and/or biochemical experiments. Often a hierarchical approach was utilized, where sequences with likelihood to form particular regular local substructures were assembled with an eye toward a specific tertiary structure [7,8]. The results yielded impressive findings, but very often, protein-like systems with structural and themodynamic properties less well defined than those of natural proteins, were obtained.

The interactions within a structured protein can have many levels of complexity making protein design approaches difficult or impractical, particularly given the large numbers of possible sequences. Proteins contain tens to thousands of amino acid residues, and even for a single sequence, many conformations of the backbone are plausible. Furthermore, just for a single backbone structure, an exponentially large numbers of side chain conformations are possible. In addition, the native conformations of proteins are largely stabilized by noncovalent forces: van der Waals, hydrophobic, electrostatic, and hydrogen bond interactions. Given the subtlety of these interactions, reliable assessments of stability with respect to unfolded states can be difficult to estimate. Finally, the large numbers of potential protein sequences can lead to combinatorial complexity in protein design, e.g., using the 20 naturally occurring amino acids, a 100-residue protein has more than 10^{130} possible sequences.

To address and overcome many of these difficulties, computational methods have been developed for the design of proteins. Most methods use as input a target structure, which can be a naturally ocurring one or one created de novo via computational modeling. To quantify interations within a given conformation and assess the compatibility between potential sequences and a selected structure, energy-based objective functions are used. The methods can then identify individual sequences or properties of sequences from a large ensemble likely to possess desirable structural and functional properties. Optimization-based methods for identifying such low-energy sequences utilize Monte Carlo simulated annealing, dead-end elimination, and genetic algorithms [9-12]. Alternatively, probabilistic methods characterize an ensemble of sequences and use ideas derived from statistical thermodynamics to estimate the site-specific probabilities of the amino acids at variable sites within the protein [12-14].

In this regard a statistical entropy-based formalism has been developed to estimate a set of amino acid probabilities for a specific backbone structure. This theory uses concepts from statistical thermodynamics to calculate the site-specific probability profiles compatibles with the given backbone structure. An effective entropy function quantifies the sequence variability consistent with the target structure. The most probable set of site-specific probabilities is determined by maximazing this effective entropy subject to centain constraints [15-17].

1.2 Molecular Dynamics Simulations, Comparative Modeling and Docking Calculations

Molecular dynamics (MD) simulations have been extensively used to estimate equilibrium and dynamic properties of proteins [18-21]. MD simulation is the computational approach to statistical mechanics and the methodology utilized is simple in principle [22]. This approach simulates motions of the system in consideration under the influence of a specific force field. Configurations in time are generated by integration of classical equations of motion [22]. By following the dynamics of a molecular system in space and in time, estimations of equilibrium and dynamic properties of the system can be obtained. Such information concerning structural and dynamic properties includes molecular geometries and energies, atomic fluctuations, rates of configurational changes, free energies, and concerted global motions [18-21,23].

In the case of proteins, all-atom MD simulations have been used to provide high-resolution information of the motions of these important biological macromolecules. These atomistic models are used to calculate statistical properties of the system that can be tested experimentally. Recent advances in computer software and hardware have increased the structural complexity and time scale simulations of the studied systems [24-28].

Many other computational tools for exploring atomic and molecular properties can complement the MD simulations approach. Some include Monte Carlo simulations, Poisson-Boltzmann anaysis, energy minimization, docking calculations, and molecular modeling.

Comparative (homology) modeling provides information of the three-dimensional protein structure, particularly for proteins that are difficult to obtain by experimental procedures such as X-ray crystallography and protein NMR structural studies. Using the comparative modeling approach, a three-dimensional model of a protein where the structure is unknown, can be created using one or more related proteins of known structure as templates. The main condition for obtaining a useful model is that the similarity between the target sequence and the template(s) sequence is detectable somehow (e.g., sequence similarity). Additionally, a correct alignment between the sequences is also essential. The comparative modeling approach is based on the notion that a small change in the protein sequence usually results in a small change in its tertiary structure [29]. In recent years template based protein modeling techniques have produced encouraging results [30-32]. Particularly relevant is the state of the art of comparative modeling for G protein-coupled receptors (GPCR), which has taken advantage of the currently increasing structural information from this membrane protein family [30,31].

The molecular docking method is extensively applied in protein systems with special relevance in rational drug design [33]. The methodology attempts to predict preferred poses or orientations of one structure (e.g., a drug-like molecule) forming an intermolecular complex with a second structure (e.g., protein receptor). Information about the preferred orientations of the structure in an intermolecular complex may be used to estimate the strength of association or binding affinity of the structures. Most docking algorithms can generate a large number of possible associated structures, thus the technique requires a way to discriminate intermolecular structures in order to identifiy those of most interest. In a nutshell, we can state that the docking problem is related with the generation and assessment of reasonable structures of intermolecular complexes [34-36]. Successful studies relevant to the work in this thesis include GPCRs, potassium channels, protein kinases, protein tyrosine phosphatase, integrin $\alpha_V\beta_3$, and G proteins [37].

1.3 Thesis Overview

Although the entropy-based formalism for computational protein design was the main technique used during most of the studies described herein, it was complemented by other computational techniques including comparative modeling, docking calculations, and specially all-atom MD simulations.

First in chapter 2, a detailed description of the entropy-based formalism used for computationally design of proteins is described. A description of the determination of the site-specific amino acid probabilities and the different constraints that modulate the sequence design are provided.

In chapter 3, the entropy-based protein design method was applied to redesign the transmembrane domain of the α subunit of the nicotinic acetylcholine receptor (nAChR). The hydrophobic protein exterior was redesigned so as to have exterior residues that are consistent with each other and with those expected on the surface of a water-soluble protein. The NMR structure of the computationally designed water-soluble structure (WSA) was solved and findings regarding the structure are discussed. Interactions with anesthetics and lipids are also analyzed. Preservation of binding sites for anesthetics and lipids in the water-soluble variant are discussed with regard to experimental findings in membrane-soluble proteins.

The structure of the human μ opioid receptor (a G protein-coupled receptor, GPCR) was built by comparative modeling techniques using recent structural information (Chapter 4).

Previous mutagenesis findings are discussed in the context of the structure. Assessment of the benefits of using different number of templates in the construction of the models is provided. Docking calculations of a typical antagonist (naloxone) and agonist (morphine) are described. Lastly, predictions of the agonist-induced activation are suggested.

Chapter 5 describes the design, experimental evaluation and MD simulations of stromal cell–derived factor–1α (SDF–1α) and simplified analogs toward the generation of effective therapeutic agents to treat inadequate blood-supply (ischemia) related heart disease. Analogs of SDF–1α were computationally modeled to simplify the protein structure. The simplified analogs were experimentally characterized in cell-based assays. A simplified analog where the central region was replaced by a two-proline linker was further characterized in a murine model of ischemic heart failure. All-atom MD simulations were carried out in SDF–1α and the simplified analogs at physiological conditions of temperature and pressure. For the simplified analogs, structural insights related with the different “success” in mimicking the beneficial therapeutic properties of SDF–1α are discussed.

Lastly, MD simulations were carried out to analyze the properties of a gyroscope-inspired molecule (Chapter 6). From NMR experiments, the structural dynamics of the novel gyroscope-inspired tribenzylamine hemicryptophane molecule indicated rigidity in one of the components (CTV stator). Also, a hindered motion of the three rotators was observed at room temperature. MD simulations showed close agreement with these experimental findings. Comparison with experimental results corroborates the utility of MD simulations in providing atomic details of molecular phenomena. Details about the rotation energy barrier and the correlation of the three rotators are discussed.

1.4 References

1. Creighton TE: *Proteins : Structures and Molecular Properties* edn 2nd. New York: W.H. Freeman; 1993.
2. Anfinsen CB: **PRINCIPLES THAT GOVERN FOLDING OF PROTEIN CHAINS**. *Science* 1973, **181**:223-230.
3. Pain RH: *Mechanisms of protein folding*. Oxford ; New York: IRL Press at Oxford University Press; 1994.
4. Pabo C: **Molecular Technology - Designing Proteins and Peptides**. *Nature* 1983, **301**:200-200.
5. Yue K, Dill KA: **Inverse Protein Folding Problem - Designing Polymer Sequences**. *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89**:4163-4167.
6. Degrado WF: **Design of Peptides and Proteins**. *Advances in Protein Chemistry* 1988, **39**:51-124.
7. Bryson JW, Betz SF, Lu HS, Suich DJ, Zhou HXX, Oneil KT, Degrado WF: **Protein Design - a Hierarchical Approach**. *Science* 1995, **270**:935-941.
8. Degrado WF, Summa CM, Pavone V, Nastri F, Lombardi A: **De novo design and structural characterization of proteins and metalloproteins**. *Annual Review of Biochemistry* 1999, **68**:779-819.
9. Samish I, MacDermaid CM, Perez-Aguilar JM, Saven JG: **Theoretical and Computational Protein Design**. *Annual Review of Physical Chemistry, Vol 62* 2011, **62**:129-149.
10. Floudas CA, Fung HK, McAllister SR, Monnigmann M, Rajgaria R: **Advances in protein structure prediction and de novo protein design: A review**. *Chemical Engineering Science* 2006, **61**:966-988.
11. Kang SG, Saven JG: **Computational protein design: structure, function and combinatorial diversity**. *Current Opinion in Chemical Biology* 2007, **11**:329-334.

12. Saven JG: **Computational protein design: engineering molecular diversity, nonnatural enzymes, nonbiological cofactor complexes, and membrane proteins**. *Current Opinion in Chemical Biology* 2011, **15**:452-457.

13. Park S, Kono H, Wang W, Boder ET, Saven JG: **Progress in the development and application of computational methods for probabilistic protein design**. *Computers & Chemical Engineering* 2005, **29**:407-421.

14. Yang X, Saven JG: **Computational methods for protein design and protein sequence variability: biased Monte Carlo and replica exchange**. *Chemical Physics Letters* 2005, **401**:205-210.

15. Saven JG, Wolynes PG: **Statistical mechanics of the combinatorial synthesis and analysis of folding macromolecules**. *Journal of Physical Chemistry B* 1997, **101**:8375-8389.

16. Zou JM, Saven JG: **Statistical theory of combinatorial libraries of folding proteins: Energetic discrimination of a target structure**. *Journal of Molecular Biology* 2000, **296**:281-294.

17. Kono H, Saven JG: **Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure**. *Journal of Molecular Biology* 2001, **306**:607-628.

18. Dodson GG, Lane DP, Verma CS: **Molecular simulations of protein dynamics: New windows on mechanisms in biology**. *Embo Reports* 2008, **9**:144-150.

19. Durrant JD, McCammon JA: **Molecular dynamics simulations and drug discovery**. *Bmc Biology* 2011, **9**.

20. Mcgeagh JD, Ranaghan KE, Mulholland AJ: **Protein dynamics and enzyme catalysis: Insights from simulations**. *Biochimica Et Biophysica Acta-Proteins and Proteomics* 2011, **1814**:1077-1092.

21. Grossfield A: **Recent progress in the study of G protein-coupled receptors with molecular dynamics computer simulations**. *Biochimica Et Biophysica Acta-Biomembranes* 2011, **1808**:1868-1878.

22. Allen MP, Tildesley DJ: *Computer simulation of liquids*. Oxford [England] New York: Clarendon Press ; Oxford University Press; 1987.

23. Stansfeld PJ, Sansom MSP: **Molecular Simulation Approaches to Membrane Proteins**. *Structure* 2011, **19**:1562-1572.

24. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE: **How Fast-Folding Proteins Fold**. *Science* 2011, **334**:517-520.

25. Gumbart J, Chipot C, Schulten K: **Free-energy cost for translocon-assisted insertion of membrane proteins**. *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**:3596-3601.

26. Dror RO, Pan AC, Arlow DH, Borhani DW, Maragakis P, Shan YB, Xu HF, Shaw DE: **Pathway and mechanism of drug binding to G-protein-coupled receptors**. *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**:13118-13123.

27. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan YB, et al.: **Atomic-Level Characterization of the Structural Dynamics of Proteins**. *Science* 2010, **330**:341-346.

28. Sgourakis NG, Garcia AE: **The Membrane Complex between Transducin and Dark-State Rhodopsin Exhibits Large-Amplitude Interface Dynamics on the Sub-Microsecond Timescale: Insights from All-Atom MD Simulations**. *Journal of Molecular Biology* 2010, **398**:161-173.

29. Chothia C, Lesk AM: **The Relation between the Divergence of Sequence and Structure in Proteins**. *Embo Journal* 1986, **5**:823-826.

30. Michino M, Abola E, Brooks CL, Dixon JS, Moult J, Stevens RC, Participants GD: **Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008**. *Nature Reviews Drug Discovery* 2009, **8**:455-463.

31. Kufareva I, Rueda M, Katritch V, Stevens RC, Abagyan R, Participants GD: **Status of GPCR Modeling and Docking as Reflected by Community-wide GPCR Dock 2010 Assessment**. *Structure* 2011, **19**:1108-1126.

32. Mariani V, Kiefer F, Schmidt T, Haas J, Schwede T: **Assessment of templete based protein structure predictions in CASP9**. *Proteins: Structure, Function, and Bioinformatics* 2011, **79**:37-58.

33. Leach AR, Shoichet BK, Peishoff CE: **Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps**. *Journal of Medicinal Chemistry* 2006, **49**:5851-5855.

34. Blaney JM, Scott DJ: **A good ligand is hard to find: Automated docking methods**. *Perspectives in Drug Discovery and Design* 1993, **1**:301-319.

35. Cavasotto CN, Phatak SS: **Homology modeling in drug discovery: current trends and applications**. *Drug Discovery Today* 2009, **14**:676-683.

36. Sun H, Scott DO: **Structure-based Drug Metabolism Predictions for Drug Design**. *Chemical Biology & Drug Design* 2010, **75**:3-17.

37. Villoutreix BO, Eudes R, Miteva MA: **Structure-Based Virtual Ligand Screening: Recent Success Stories**. *Combinatorial Chemistry & High Throughput Screening* 2009, **12**:1000-1016.

2 Probabilistic Theory of Computational Protein Design

2.1 Statistical Theory of Sequence Ensembles

Identifying the properties of sequences compatible with a predetermined backbone structure would provide ways to probe the determinants of protein folding and identify folding amino acids sequences [1,2]. A probabilistic theory has been previously introduced to explore and characterize sequence space for a target structure without explicitly tabulating all sequences [3,4]. The statistical entropy-based formalism has been developed to determine a set of amino acid probabilities for a chosen backbone structure. This approach laverages concepts from statistical thermodynamics to estimate site-specific probability profiles compatible with the given backbone structure [1].

To adress the energetic sequence compatibility with regard to the backbone structure, the methodology includes several scoring functions. In addition to an atom based potential (physically derived), knowledge based potentials can be included to modulate centain properties of the system such as secondary structure preference and hydrophobicity. The sequence conformational space can be readily explored by the method and as a consequence, large systems (> 100 residues; for 100-residue protein, $20^{100} \approx 10^{130}$ possible sequences) can be considered in the protein design calculations. The method yields the probabilities of each amino acid at each of the positions of interest in a given protein structure and not only a specific sequence (oftentimes, the minimum energy solution) [5]. The generality of the method allows different prerequisites for foldable sequences or arbitrary constraints such as patterning of hydrophobic residues to be included.

In specifying the sequence identity and energetic conformations subject to certain constraints for a specific folded state, the probability of amino acid α at position i and with a side chain in a conformation state defined by $r(\alpha)$, is denoted by $w_i(\alpha,r(\alpha))$. The total sequence-conformational entropy function S that quantifies the variability of sequences consistent with a particular target structure is defined by equation 2-1.

$$S = -\sum_{i} \sum_{\alpha,r(\alpha)} w_i\left(\alpha,r(\alpha)\right) \ln w_i\left(\alpha,r(\alpha)\right) \tag{2-1}$$

The expression considers each of the sequence position i and all available amino acid identities α. For each amino acid identity α, the expression also considers all the possible rotamer states $r(\alpha)$ [6]. The residues state probabilities are obtained by maximizing this effective entropy function subject to different constraints. The maximization is done using the Lagrange multipliers method [7]. The variational functional of the set of probabilities $w_i(\alpha,r(\alpha))$ is defined by equation 2-2.

$$V = S - \lambda_1 f_1 - \lambda_2 f_2 - \lambda_3 f_3 - \cdots \tag{2-2}$$

The constraint functions f_k specify global and local features of the sequence-structure space and are in general, function of the probabilities $w_i(\alpha,r(\alpha))$. λ_k are the respective Lagrange multipliers. To identify the probabilities consistent with particular value(s) of the constraint(s) f_k, the constraint function is restrained to have a particular value f_k^0.

$$f_k^0 = f_k\left(\{w_i(\alpha,r(\alpha))\}\right) \tag{2-3}$$

The large set of coupled nonlinear equations that determine the set of state probabilities $\{w_i(\alpha,r(\alpha))\}$ and the Lagrange multiplier λ_k associated with the contraints are defined by equations 2-4 and 2-5.

$$0 = \frac{\partial V}{\partial w_i\left(\alpha, r(\alpha)\right)} \qquad (2\text{-}4)$$

$$0 = \frac{\partial V}{\partial \lambda_k} \qquad (2\text{-}5)$$

2.2 Constraints

2.2.1 Energy Functions

Conformational energy function E that quantifies sequence-structure compatibility is determined using a molecular mechanics force field [8]. This conformational energy is composed of non-bonding interactions including van der Waals (using a Lennard-Jones 12-6 potential), electrostatics with a distance dependent dielectric constant ($\varepsilon = 4\varepsilon_0 r_{ij}$ where r_{ij} is the interatomic distance), and a modified angular-dependent hydrogen bond term [9] (additionally, a torsional energy term for the side chains is sometimes included as one body term, see equation 2-6). For a particular sequence of N sites $(\alpha_1,\ldots,\alpha_N)$ where the side chain conformational states of the each amino acid are $(r(\alpha_1),\ldots, r(\alpha_N))$, the conformational energy E can be expressed as a one-body and a pairwise two-body terms, equation 2-6.

$$E = \sum_i \varepsilon_i^{(1)}\left(\alpha, r(\alpha)\right) + \sum_{i<j} \gamma_{ij}^{(2)}\left(\alpha, r(\alpha); \alpha', r(\alpha')\right) \qquad (2\text{-}6)$$

The first term contains the one-body contribution $\varepsilon_i^{(1)}(\alpha,r(\alpha))$ and includes the interaction between backbone and amino acid side chains. Additionally, it contains the reference energy that accounts for the unfolded state contribution, see section 2.2.2. The second term contains the two-body contribution $\gamma_{ij}^{(1)}(\alpha,r(\alpha);\alpha',r(\alpha'))$ and includes all the interaction energies between amino acid side chains $r(\alpha)$ and $r(\alpha')$ at site i and j, that is two different rotamer states at two different sites in the structure.

Within the context of the statistical theory, for a given set of sequences, the assumption that the fluctuations in the conformational energy around the main value (due to variability of sequence) are small is reasonable. The conformational energy can be writen as equation 2-7.

$$E \approx \sum_{i} \sum_{\alpha,r(\alpha)} \varepsilon_i^{(1)}(\alpha,r(\alpha))w_i(\alpha,r(\alpha)) + \sum_{i<j} \sum_{\alpha,r(\alpha)} \sum_{\alpha',r(\alpha')} \gamma_{ij}^{(2)}(\alpha,r(\alpha);\alpha',r(\alpha'))w_i(\alpha,r(\alpha))w_j(\alpha',r(\alpha')) \tag{2-7}$$

2.2.2 Reference Energy

The paradigm in protein design is to optimize the energy of a particular sequence in the folded structure. In order to discriminate among a large number of possibilities, information about the energetics of the folded and unfolded states must be considered [2]. The discrimination of different sequences is carried out relative to the energy of the ensemble of unfolded states [2,10]. The calculation of the energy of the ensemble states is very complex even for small systems. To address this difficulty, a reference energy γ_{ref} is introduced to account for the energetics of the unfolded states, equation 2-8. Using the structure: N–acetyl–(α)–N'–methylamide, the Helmholtz free energy is calculated for each amino acid (α) over different possible rotamer states $r(\alpha)$ and backbone degrees of freedom (ϕ, ψ). In particular, the calculation of the partition function z_{ref} at a given temperature T for each of the amino acids involves the sum of the conformational energy over different rotamers states and variations of the backbone degrees of freedom (ϕ, ψ). The different rotamer states are those from a backbone dependent rotamer library [11]. The variations of the ϕ and ψ angles involved increments in 10° from 0 to 2π, equation 2-9. The results for the reference energy γ_{ref} for the different amino acid residues are expressed with respect to the value for glycine.

$$\gamma_{ref}(\alpha,T) = -k_B T \ln \frac{z_{ref}(\alpha,T)}{z_{ref}(Gly,T)} \tag{2-8}$$

$$z_{ref}(\alpha,T) = \sum_{\phi,\psi} \sum_{r(\alpha)} \exp\left(- \frac{\varepsilon_{ref}(\phi,\psi,r(\alpha))}{k_B T} \right) \tag{2-9}$$

Since the reference energy is just a function of the amino acid type at each site, this term that accounts for the contributions of the unfolded states is introduced directly in the effective one-body term in equation 2-7. Consequently, the energy contraint that involves interatomic interaction can be rewritten as equation 2-10.

$$\begin{aligned} E \approx & \sum_{i} \sum_{\alpha,r(\alpha)} \left(\varepsilon_i^{(1)}(\alpha,r(\alpha)) - \gamma_{ref}(\alpha,T) \right) w_i(\alpha,r(\alpha)) \\ & + \sum_{i<j} \sum_{\alpha,r(\alpha)} \sum_{\alpha',r(\alpha')} \gamma_{ij}^{(2)}(\alpha,r(\alpha);\alpha',r(\alpha')) w_i(\alpha,r(\alpha)) w_j(\alpha',r(\alpha')) \end{aligned} \tag{2-10}$$

2.2.3 Environmental Energy

To account for solvation effects and for the tendency of different amino acids to be exposed to or sequestered from water (hydrophobicity), an additional contraint was introduced, herein environmental energy (E_{env}) [4]. This energy term is based on the local density of C_β atoms of each residue in the fixed protein scaffold. Generally in a folded structure, hydrophobic residues tend to be sequestered from the solvent and as a consequence, surrounded by a large number of neighbors. The number of C_β atoms in the vicinity of these buried residues is relatively large. On the other hand, hydrophilic residues are prone to be located at the surface positions where they are highly exposed to water molecules. In this case, the amount of neighbors and thus, the number of C_β atoms in their vicinity, is relatively low. The choice of C_β density is convenient not only because this quantity is invariant for a specific fixed backbone structure but also because the implementation is complatible and consistent with the statistical implementation described in section 2.1 [4]. The effective local environmental energy is defined by equation 2-11.

$$\varepsilon_{env}(\alpha, r(\alpha); \rho) = -k_B T \ln \frac{f(\alpha, \rho)}{f(\alpha) f(\rho)} \qquad \text{(2-11)}$$

Where ε_{env} is the effective local environmental energy that depends on the amino acid type α with particular rotamer state $r(\alpha)$ and the specific C_β density environment ρ. $f(\alpha,\rho)$ is the fraction of times a local C_β density ρ, is observed for the amino acid type α. $f(\alpha)$ is the fraction of times a local density ρ is observed in the globular protein training set used for the parametrization. $f(\rho)$ is the fraction of times a local density ρ is observed irrespective of the amino acids type. The protein-training database was composed of 500 representative globular protein structures (http://www.cbrc.jp/pdbreprdb-cgi/reprdb_query.pl), see details in reference [4]. The local C_β density ρ (equation 2-12) is defined as the density of C_β atoms within the free volume of a sphere of radius R_c centered at the center of mass of each side chain. The free volume is referred as the volume not excluded by the side chain and its respective peptide main chain.

$$\rho(\alpha) = \frac{n_\beta\left(C_\beta, R\right)}{\frac{4}{3}\pi R_c^3 - \left\langle V_{access}(\alpha) \right\rangle} \qquad \text{(2-12)}$$

Here n_β is the number of C_β atoms within a specified volume of radius R and $\langle V_{access}(\alpha)\rangle$ is the average volume not excluded by the side chain of α, where the average is taken over all possible rotamer states $r(\alpha)$. The average value of the C_β density $\langle\rho(\alpha)\rangle$ for each amino acid was found to correlate well other scales of hydrophobicity [4,12].

Within the context of the statistical theory, the environmental energy is defined by equation 2-13. Noteworthy, for a fixed backbone, $E_{env}(\alpha,r(\alpha))$, which depends on both the amino acid type and the rotamer state, is expressed as one-body term. The calculation of the environmental energy in the globular protein-training set correlated well with the length of the protein (number of residues). During the sequence probability calculation, the target E^0_{env} is constrained to a value in concordance with the linear relationship found in the globular protein set between the value of the environmental energy and the number of residues of the system. The ensemble environmental energy E_{env} is optimized so as to satisfy the target value of the environmental constraint E^0_{env}.

$$E_{env}^{0} = E_{env} = \sum_{i}\sum_{\alpha,r(\alpha)}\varepsilon_{env}\big(\alpha,r(\alpha)\big)w\big(\alpha,r(\alpha)\big) \qquad (2\text{-}13)$$

2.2.4 Diversity Entropy Potential

During the framework of the statistical method in protein design, the selection of a final sequence is usually carried out by using the most probable amino acids at each of the target positions. Sometimes, different criteria are used to obtain the final consensus sequence. In the case of buried positions, the environment exerts restraints that tend to bias the preference for some amino acids. In these positions, a distinctive and easy to interpret probability distribution is frequently observed. On the contrary, in the case of highly exposed positions, the solutions obtained in the statistical calculations frequently display unbiased distributions for the amino acid probabilities. Polar or charged amino acids exhibit very similar probabilities and are slightly preferred than non-polar amino acids. Depending upon the purpose of the design, sometimes a more diverse sequence is desired as in the case of NMR experiments where such diversity should yield dispersion in the NMR spectrum and facilitate structure determination. To increase the amino acid diversity, an additional constraint consistent with the statistical methodology was introduced. This constraint, denominated diversity entropy potential S_I, was implemented as an inverse participation ratio.

$$S_I = \frac{1}{m_{eff}} = \sum_{\alpha} f^2(\alpha) \qquad (2\text{-}14)$$

$$f(\alpha) = \frac{\sum_{i=1}^{n} w_i\big(\alpha, r(\alpha)\big)}{n} \qquad (2\text{-}15)$$

Here, $f(\alpha)$ is the frequency with which amino acid α appears among the sequences and is defined as the probability of amino acid α summed over the n targeted positions (targeted positions are usually all the positions selected to be mutated during the design). The number of possible amino acids, m, is m = 20 if all the naturally occurring amino acids are permitted. The effective number of amino acids that appear at the targeted sites, m_{eff}, can be defined using the calculated amino acid probabilities; m_{eff} is bounded $1 \le m_{eff} \le m$ and it represents an average over the ensemble of sequences. For a homopolymer where only one amino acid is used to construct the sequence, m_{eff} = 1; if all 20 amino acids appear with equally frequency, m_{eff} = 20. Thus m_{eff} may be viewed as the effective number of amino acids used to construct the sequences.

2.3 Site-Specific Amino Acid Probabilities

In solving the coupled nonlinear equations that maximize the conformational entropy function subject to certain constraints, the site-specific probabilities and the associated Langrange multipliers to those constraints are determined.

$$w_i(\alpha) = \sum_{r(\alpha)} w_i\left(\alpha, r(\alpha)\right) \qquad \text{(2-16)}$$

The set of probabilities are obtained at different effective temperatures β^{-1}, where β is the Lagrange multiplier associated with the conformational energy constraint. At β equal to 0.5 mol/kcal the sequence properties are found to be robust with respect to slight changes in the backbone structure [4].

2.4 References

1. Saven JG, Wolynes PG: **Statistical mechanics of the combinatorial synthesis and analysis of folding macromolecules**. *Journal of Physical Chemistry B* 1997, **101**:8375-8389.
2. Saven JG: **Designing protein energy landscapes**. *Chemical Reviews* 2001, **101**:3113-3130.
3. Zou JM, Saven JG: **Statistical theory of combinatorial libraries of folding proteins: Energetic discrimination of a target structure**. *Journal of Molecular Biology* 2000, **296**:281-294.
4. Kono H, Saven JG: **Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure**. *Journal of Molecular Biology* 2001, **306**:607-628.

5. Samish I, MacDermaid CM, Perez-Aguilar JM, Saven JG: **Theoretical and Computational Protein Design**. *Annual Review of Physical Chemistry, Vol 62* 2011, **62**:129-149.
6. Dunbrack RL: **Rotamer libraries in the 21(st) century**. *Current Opinion in Structural Biology* 2002, **12**:431-440.
7. McQuarrie DA: **Statistical Mechanics**. *Sausalito CA: University Science Books* 2000.
8. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P: **A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins**. *Journal of the American Chemical Society* 1984, **106**:765-784.
9. Kono H, Doi J: **A new method for side-chain conformation prediction using a Hopfield network and reproduced rotamers**. *Journal of Computational Chemistry* 1996, **17**:1667-1683.
10. Wernisch L, Hery S, Wodak SJ: **Automatic protein design with all atom force-fields by exact and heuristic optimization**. *Journal of Molecular Biology* 2000, **301**:713-736.
11. Dunbrack RL, Cohen FE: **Bayesian statistical analysis of protein side-chain rotamer preferences**. *Protein Science* 1997, **6**:1661-1681.
12. Fauchere JL, Pliska V: **Hydrophobic Parameters-Pi of Amino-Acid Side-Chains from the Partitioning of N-Acetyl-Amino-Acid Amides**. *European Journal of Medicinal Chemistry* 1983, **18**:369-375.

3 NMR Structure and Dynamics of a Computationally Designed Water-Solubilized Transmembrane Domain of the Nicotinic Acethylcholine Receptor (nAChR) [1, 2]

3.1 Introduction

Nicotinic acetylcholine receptors (AChRs) belong to a superfamily of neurotransmitter-gated ion channels, which also include glycine (Gly), γ-amino-butyric acid type A ($GABA_A$), and serotonin ($5\text{-}HT_3$) receptors. These receptors, composed of five homologous subunits in a pentameric assembly, mediate fast synaptic transmissions. All subunits comprise of a large extracellular domain, four transmembrane (TM) domains, and a long intracellular loop linking the third and fourth TM domains [1,2]. Mutagenesis and functional investigations have showed that binding of at least two agonists at the interfaces of the extracellular domains triggers the channel opening, allowing ions to pass through a TM pore surrounded by the TM2 helices from the five subunits.

---

[1] This project was a collaboration with the group of Professor Roderick G. Eckenhoff at the Department of Anesthesiology and Critical Care Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104 and the group of Professor Yan Xu at the Department of Anesthesiology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15260.

[2] Adapted from Tanxing Cui, David Mowrey, Vasyl Bondarenko, Tommy Tillman, Dejian Ma, Elizabeth Landrum, Jose Manuel Perez-Aguilar, Jing He, Wei Wang, Jeffery G. Saven, Roderic G. Eckenhoff, Pei Tang, Yan Xu. *Biochimica et Biophysica Acta – Biomembranes* **2012**, 1818, 617-626.

The most studied member of this receptor superfamily is the highly abundant nAChR found in the *Torpedo* electric organs. The *Torpedo* nAChR, consisting of two $\alpha 1$ subunits and one each of $\beta 1$, γ, and δ subunits, is homologous to the mammalian muscle-type nAChR responsible for the rapid signaling at the neuromuscular junction. Although no atomic resolution structure is yet available for nAChRs, a 4-Å-resolution structure model has been determined by the pioneering cryo-electron microscopy (cryo-EM) work on the *Torpedo* nAChR [3,4].

More recently, high-resolution structural insights are gained from several crystal or NMR structures of related proteins, including an isolated extracellular domain of the nAChR $\alpha 1$ subunit bound to α-bungarotoxin [5], isolated four TM domains of the nAChR $\beta 2$ subunit [6], the snail acetylcholine binding protein [7] which resembles the extracellular domain of the nAChR, and two prokaryotic pentameric ligand-gated ion channels from the bacterium *Erwinia chrysanthemi* (ELIC) [8] and the cyanobacterium *Gloebacter violaceus* (GLIC) [9,10]. The overall architecture of the extracellular domains is similar among the currently available structures. The TM domains, however, exhibit significant variation in high-resolution details. The discrepancy among the TM domain structures may result from the difference in the membrane mimetic environments used for the structure determination, or from the difference in the ion-conducting states of the TM segments. It is worth noting that the 4-Å-resolution cryo-EM structure of the muscle-type nAChR shows four straight long helices in the TM domains. In contrast, the GLIC and ELIC crystal structures and the nAChR $\beta 2$ NMR structure all exhibit relatively short and curved helical structures. The cryo-EM nAChR structure is presumably in the closed-channel state, whereas GLIC is either in a desensitized state or in a non-physiological artificial state due to the unexpected presence of detergent molecules within the aqueous pore. The ELIC crystal structure is thought to be in a collapsed state. Potential interactions with cholesterols, which are present in eukaryotic membrane but absent in prokaryotic membrane, can add another layer of complexity in TM domain structure determination.

The uncertainty in the TM domain structures reflects the technical difficulties associated with the structure determination of TM channel proteins. Crystallization trials involve the use of nonnative detergents that may or may not support the functional fold of the TM domains. The success in the expression, purification, and crystallization of prokaryotic channels such as ELIC and GLIC does not readily translate into practice for mammalian channels, which are difficult to overexpress in suitable quantities for structural studies. Because of their large hydrophobic surface in the TM segments and the need for posttranslational modifications, mammalian channel proteins are also prone to aggregation, rendering the standard solution-phase purification methods ineffective. Most biophysical studies involve dispersing membrane proteins in aqueous media, usually using detergents, lipids, auxiliary proteins, and other membrane-mimetic mixtures. Obtaining conditions to achieve monodispersity for X-ray crystallography or NMR studies remains a touchy and exceedingly time consuming process.

An alternate, systematic approach is to redesign membrane proteins to remove the exposed hydrophobic exterior. In essence, TM proteins are transformed into soluble ones while retaining important structural and functional features. Using computational approaches, water-soluble variants of integral membrane proteins have been designed to facilitate structural studies. Recently, a computationally designed water-soluble variant of the bacterial potassium ion channel KcsA has been accomplished [11,12] and its high-resolution structure has been elucidated by NMR [13]. The NMR structure of one of the three designed sequences of water-solubilized KcsA (WSK3) is in striking agreement with the crystal structure of the wild-type protein [13,14]. These findings emphasize the promise of developing water-soluble variants of membrane proteins suitable for biophysical studies and potentially for structure-based drug discovery.

In this study, we present a similar approach for the TM domains of nAChR $\alpha 1$ subunit. A water-soluble variant of the TM domains was computationally designed, resulting in the replacement of 23 hydrophobic residues with hydrophilic ones at the lipid-exposed surface. The resulting protein, called water-soluble acetylcholine receptor channel (WSA), was experimentally expressed and purified, and its structures determined using high-resolution NMR. Because the nAChRs are potential molecular targets for general anesthetics [15-19], we also determined the binding of two representative anesthetics, azi-isoflurane and azi-propofol, to the designed protein using photoaffinity labeling. We found that the designed protein retained an overall four-helix bundle topology and overlapped remarkably well with the crystal structure of the GLIC TM domains. A common binding site was identified for azi-isoflurane and azi-propofol in the general vicinity of a propofol binding site found in the crystal structure of anesthetic-bound GLIC. These results provide additional experimental evidence for the general applicability of the water solubilization approach to high-resolution structural investigation of membrane proteins.

3.2 Methods

3.2.1 Computational Sequence Design for Water Solubilization

The design principle involves the identification and redesign of exposed hydrophobic residues in the TM domains in order to transform a membrane protein into a completely or partially soluble protein while maintaining the structure- and function-related properties. The redesign of the nAChR $\alpha 1$ subunit was based on the 4-Å-resolution cryo-EM structure model as a template (PDB accession code: 1OED; Figure 3.1a) [3,4]. The exposed residues were selected as those with more than 40% solvent exposure (Figure 3.1b) according to GETAREA1.1 [20]. Residues targeted for mutation were both hydrophobic (AFILMVW) and expected to reside in the TM region [3]. Hydrophobic residues at the interfaces between helices within the subunit were retained as in the wild type protein.

The site-specific probabilities of the amino acids at the potential mutation sites were calculated as described previously [11,21-25]. For each of the sites, all 20 amino acids were permitted while the remaining residues were constrained as wild type. Side chain conformations were restricted to those from a library of frequently observed conformations (rotamer states) [26] with at most 10 rotamer states per amino acid. Two energy functions, averaged over the sequence of the entire protein, were constrained to characterize the ensemble of soluble variants. The first energy term accounts for the inter-atomic interactions using a molecular mechanics force field (the AMBER force field) [27]. The second term quantifies the hydrophobic potential (environmental or solvation energy) that was constrained to an averaged value expected for a soluble protein having the size of the TM domains of the nAChR α1 subunit [11].

The computation for site-specific probabilities of amino acids at each mutation position was carried out in two stages. The first stage used the most probable amino acids at each of the targeted positions. At the second stage, only the sites where one amino acid was highly favored (probability of an amino acid exceeding 0.8) were constrained to that most probable amino acid. At the rest of the positions, an additional constraint was imposed to increase the diversity of different amino acids that appear in the final sequence. Amino acid diversity is desirable to avoid spectral over-crowding in the NMR spectra, thereby facilitating spectral assignment and structure determination. The sequence diversity constraint was implemented as an inverse participation ratio:

$$\frac{1}{m_{eff}} = \sum_{\alpha=1}^{m} f^2(\alpha) \qquad (3\text{-}1)$$

$$f(\alpha) = \frac{\sum_{i=1}^{n} w_i(\alpha)}{n} \qquad (3\text{-}2)$$

where m is the number of possible amino acids, e.g., m = 20 if all the naturally occurring amino acids are permitted. The function $f(\alpha)$ is the frequency with which amino acid α

appears among the sequences and is defined as the probability of amino acid α summed over the n targeted positions. The effective number of amino acids that appear at the targeted sites, m_{eff}, can be defined using the calculated amino acid probabilities; m_{eff} is bounded $1 \leq m_{eff} \leq m$ and it represents an average over the ensemble of sequences. For a homopolymer where only one amino acid is used to construct the sequence, $m_{eff} = 1$; if all 20 amino acids appear with equal frequency, $m_{eff} = 20$. Thus m_{eff} may be viewed as the effective number of amino acids used to construct the sequences.

The current design focused on the TM domain without the extracellular domain and the intracellular loop. The removal of the intracellular loop separate TM4 from TM1-TM3. In order to express the designed protein as a single chain, a polyglycine linker was designed using the loop builder in MODELLER [28]. To minimize any structural biases and effects of an artificial linker on TM3 and TM4 and to take advantage of the additional length from the residues in the wild-type sequence leading to the TM1 helix and ending the TM4 helix (the C-terminus), we placed the artificial polyglycine linker between the C-terminus of helix TM4 and the N-terminus of helix TM1 (Figure 3.1c). This arrangement preserves the orientations of individual TM domains without restricting the relative spatial positions of TM3 and TM4, as in the topology of the original cryo-EM template.

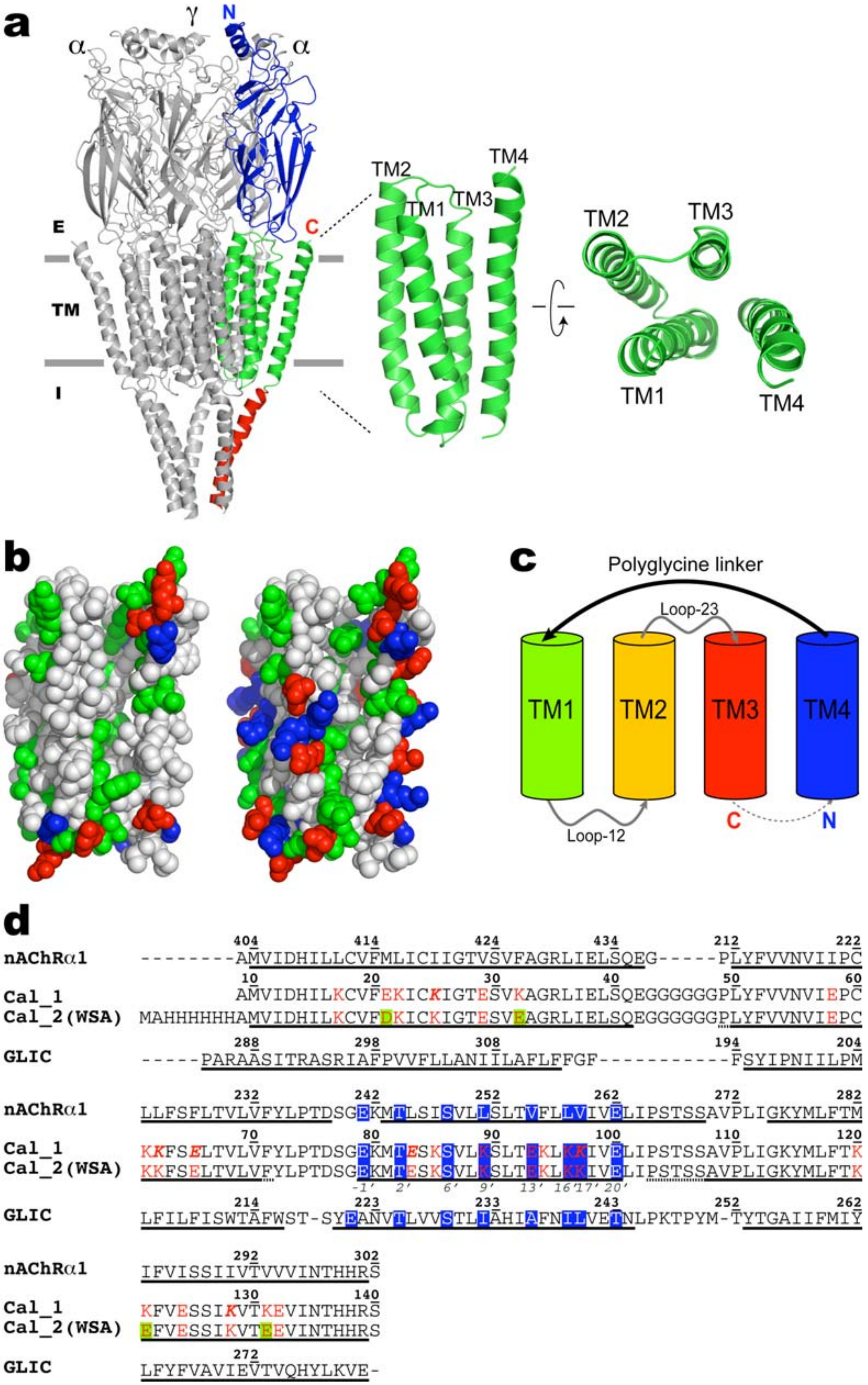

a
γ
N
α
α
E
TM
I
C
TM2
TM1
TM3
TM4
TM2
TM3
TM1
TM4
b
c
Polyglycine linker
Loop-23
TM1
TM2
TM3
TM4
Loop-12
C
N
d
404 414 424 434 212 222
nAChRα1 ---------AMVIDHILLCVFMLICIIGTVSVFAGRLIELSQEG-----PLYFVVNVIIPC
10 20 30 40 50 60
Cal_1 AMVIDHILKCVFEKICKIGTESVKAGRLIELSQEGGGGGGPLYFVVNVIEPC
Cal_2(WSA) MAHHHHHHAMVIDHILKCVFDKICKIGTESVEAGRLIELSQEGGGGGGPLYFVVNVIEPC
288 298 308 194 204
GLIC -----PARAASITRASRIAFPVVFLLANIILAFLFFGF-----------FSYIPNIILPM
232 242 252 262 272 282
nAChRα1 LLFSFLTVLVFYLPTDSGEKMTLSISVLLSLTVFLLVIVELIPSTSSAVPLIGKYMLFTM
70 80 90 100 110 120
Cal_1 KKFSELTVLVFYLPTDSGEKMTESKSVLKSLTEKLKKIVELIPSTSSAVPLIGKYMLFTK
Cal_2(WSA) KKFSELTVLVFYLPTDSGEKMTESKSVLKSLTEKLKKIVELIPSTSSAVPLIGKYMLFTK
-1′ 2′ 6′ 9′ 13′ 16′17′ 20′
214 223 233 243 252 262
GLIC LFILFISWTAFWST-SYEANVTLVVSTLIAHIAFNILVETNLPKTPYM-TYTGAIIFMIY
292 302
nAChRα1 IFVISSIIVTVVVINTHHRS
130 140
Cal_1 KFVESSIKVTKEVINTHHRS
Cal_2(WSA) EFVESSIKVTEEVINTHHRS
272
GLIC LFYFVAVIEVTVQHYLKVE-

Figure 3.1. Sequence Redesign. (**a**) Ribbon representation of the pentameric structural template (PDB code: 2BG9) of nicotinic acetylcholine receptor (nAChR) used for the water solubilization sequence redesign. (**b**) The transmembrane domain of the α1 subunit in van der Waals representation before (left) and after (right) the computational redesign. Residues are colored by amino acid types: Green: hydrophilic (GNQSTY); white: hydrophobic (ACFILMPVW); blue: basic (HKR); and red: acidic (DE). (**c**) Cartoon representation of helix configuration in water-solubilized acetylcholine receptor (WSA). A polyglycine linker was added between TM1 (helical N-terminus) and TM4 (helical C-terminus) so that the TM domains can be expressed as a single chain. (**d**) Sequences of the transmembrane domains of nAChR α1 subunit, two calculated water-solubilization designs, and the bacterial pentameric ion channel GLIC are aligned for comparison. Water-solubilization mutations are marked in red. The most probable mutations are in italic boldface. Mutations changed by the inclusion of sequence diversity constraints are highlighted in green. Putative pore-lining residue positions are highlighted in blue. The transmembrane domain helices in the corresponding structures are underlined for comparison. Conformation heterogeneity in WSA is underlined with dashed lines.

3.3 Results

3.3.1 WSA Sequence Design

Using the cryo-EM structure model of the nAChR α1 subunit [3,4] as a template and 40% solvent accessibility surface as the criterion for exposed residues, we identified 29 hydrophobic residues within the TM domains as potential redesign targets. Of these, six were considered at the interfaces between the helices and kept unchanged: I219, M243, V261, F414, I420 and V425. The remaining 23 positions were targeted for computational redesign. They are I220, L223, L224, F227, L245, I247, L251, V255, F256, L258, V259, M282, I283, I286, I290, V293, V294, L411, M415, L416, I419, V423, and F426.

At the first stage of site-specific probability computation, the most probable amino acids at each of the 23 positions were calculated. The resulting sequence is shown in Figure 3.1d as Cal_1. This sequence was found to be cytotoxic to *E. coli*, and the resulting expression yield was low. In the second stage of redesign, only the positions with amino acid probability > 0.8 were constrained to the most probable amino acids. The following six mutations satisfied this condition: L224K, F227E, L245E, V259K, I290K and I419K. At the remaining 17 positions, sequence diversity constraints were used according to Eq. (3-1) and Eq. (3-2). With 23 variable residues, the site-specific probabilities yield m_{eff} = 4.3 in the absence of any constraints on diversity. With the mutations at the above six positions fixed, the constraint is applied so as to achieve m_{eff} = 5.0. The most probable amino acid was selected at each of the 17 variable positions, yielding sequence Cal_2 in Figure 3.1d. This sequence has a theoretical isoelectric point (pI) of = 6.39 [29]. Also displayed in Figure 3.1d is an aligned sequence for the GLIC TM domains. The sequences are 83% identical between Cal_2 and the original α1 subunit of nAChR, and 11% between Cal_2 and GLIC. The residue numbering used in this study and the corresponding sequence numbering in nAChR α1 subunit and in GLIC are shown in Figure 3.1d.

The lengths of the polyglycine linker (4 to 8 Glys) were investigated, and 50 independent calculations for each length of loop were performed using MODELLER [28]. The 5-glycine segment had the lowest effective energy (best score in MODELLER). The quality of the choice of loop length was also confirmed by energy minimization of the loop structures with NAMD [30] using the CHARMM force field [31].

3.3.2 Secondary Structure of WSA by CD and NMR

Circular dichroism (CD) measurenments of the water-soluble variant in aqueous media with 2% of lipid (1-palmitoyl-2-hydroxy-*sn*-glycero-3-[phospho-*rac*-(1-glycerol)] (LPPG)) show an α-helical content of ~50%. Although this value is significantly lower than the helical content of the cryo-EM template (~83-87%), it is in reasonable agreement with the estimation from FTIR and CD measurement of the authentic full-length nAChR [32] and GlyR [33,34]. From the NMR chemical shift index (CSI), secondary structure can be inferred. In the case of WSA, this quantity indicates four helical segment (Figure 3.2b), with ~70% residues having a Cα chemical shift consistent with an α-helical secondary structure. These results are in excellent agreement with the helical content of the transmembrane domains of GLIC. Additionally, a slow conformational exchange between two conformations for the segment ^{104}STSS^{107} was observed on the NMR timescale (Figure 3.2c).

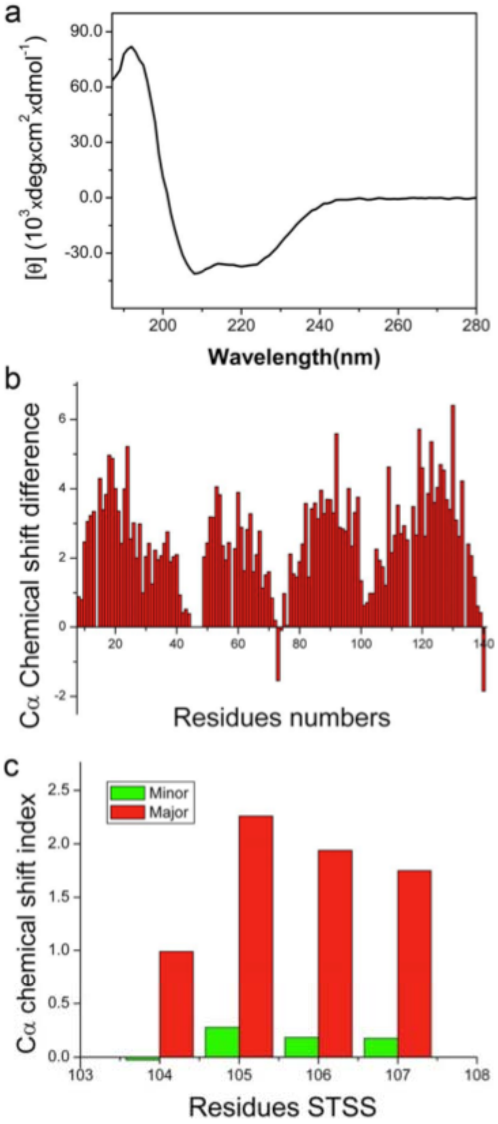


Figure 3.2. Secondary Structure Determination. (**a**) Far-UV circular dichroism (CD) spectrum of WSA in water with 2% LPPG, showing an α-helical content of ~50%. (**b**) The NMR chemical shift index (CSI), determined based on the backbone Cα resonance frequencies relative to the random coil values, is plotted as a function of residue numbers. Four helical domains are clearly defined. (**c**) Conformation heterogeneity in the segment between TM2 and TM3 is visible in the resonance peak doubling for residues ^{104}STSS^{107}. SCI suggests slow exchange between helical and non-helical conformations.

3.3.3 High-Resolution NMR Structure of WSA

The advantage of water solubilization of TM channel proteins is clearly displayed in the NMR spectra of WSA (Figure 3.3a). Not only can WSA be expressed and purified in large quantity for structural studies, the NMR spectra also showed that the water solubilization of this TM protein greatly improved monodispersity of the protein sample, allowing high-resolution NMR spectra to be acquired with homogenous peak intensities. Well-resolved 3D NMR spectra permitted spectral assignment of nearly all backbone resonances and many side chain resonances. With the exception of the segment STSS mentioned above, the majority of residues in the WSA sequence showed one set of NMR peaks, indicating a structural homogeneity. The structures of WSA with the STSS segment in the major and minor conformations were determined separately using the standard solution-state NMR methods. The 15 lowest target-function structures for each conformation are shown in Figure 3.3b (protein data bank PDB ID code 2LKG and 2LKH for the major and minor conformations, respectively), and the structure with minimal root-mean-square deviation (RMSD) from the rest in the same conformation is highlighted in the ribbon representation. The overall backbone RMSD in the four helical domains is 0.69 ± 0.13 Å and 0.93 ± 0.19 Å in the major and minor conformations, respectively. None of the structures have a distance violation > 0.5 Å or an angle violation > 5°, and 99.2% of the residues are in the most favored and additional allowed regions of the Ramachandran plot and no residues are in the disallowed region. Lastly, backbone ^{15}N R_1 and R_2 relaxation parameters suggested a monomeric form of WSA.

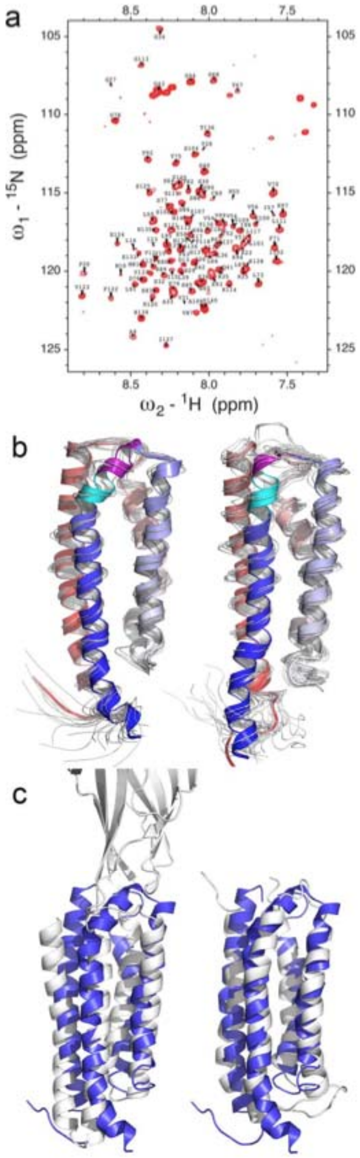


Figure 3.3. Structure Determination of WSA by High-resolution NMR. (**a**) Representative ^{1}H-^{15}N HSQC NMR spectrum of WSA acquired at 40°C on an 800-MHz spectrometer, showing well resolved resonances of the water-solubilized membrane protein. (**b**) Bundles of the 15 lowest target function structures of major (left) and minor (right) conformations. The structures with the least RMSD from the rest of the bundle are depicted in the ribbon representation, with the TM4-TM1-TM2-TM3 domains colored in a red-to-blue color scale. The STSS segment is highlighted in magenta, and Loop-23 in the Cryo-EM template, which is a part of the TM3 helix in the WSA structure, is highlighted in cyan. (**c**) Superposition of the WSA NMR structure (blue) onto the TM domains of the crystal structure of GLIC (left, white) and the NMR structure of nAChR β2 subunit (right, white), which was solved in hexafluoroisopropanol [6].

3.3.4 Anesthetic Binding to WSA

Using photoaffinity labeling, we investigated potential anesthetic binding to WSA. Peptides accounting for 48% of the sequence were detected for the azi-isoflurane labeled sample, and a single adduction site was found at V-31. Similarly, peptides accounting for 67% of the WSA sequence were detected in the aziPm labeled sample, and the same residue (V-31) was found to be the only modified site.

3.4 Discussion

3.4.1 Structure Characteristics of WSA

We have solved the structure of a water-solubilized analogue of the TM domains of nAChR α1 subunit using solution-state NMR spectroscopy. The protein was designed to be monomeric in solution, permitting secondary and tertiary structures to be determined at the atomic resolution. For clarity, we focus our discussion of the WSA structures with the STSS segment in its major conformation. The structure shows the expected four-helix bundle fold but with four unequal helix lengths. TM1 and TM2 are both six helix-turns long, shorter than TM3 (9 helix turns) and TM4 (8 helix turns). In comparison, the cryo-EM structure of *Torpedo* nAChR, which was used as the template in the computational sequence redesign, has four straight helices of roughly equal lengths (8 helix turns for TM1, TM2, and TM3, and 10 helix turns for TM4). The residues belonging to the helices as determined in the cryo-EM structure and the WSA NMR structure, along with those in the crystal structure of GLIC, are underlined in the sequences in Figure 3.1d for comparison. A superposition of WSA structure onto the monomer structure of GLIC is depicted in Figure 3.3c. Although WSA has ~83% sequence identity to the wild-type nAChR and only 11% identity to the GLIC TM domains, the high-resolution NMR structure of WSA agrees more with that of GLIC than with the cryo-EM template. This is rather unexpected. Several contributing factors can be considered. First, the experimental conditions under which the structures were determined were very different. The cryo-EM structure of nAChR was determined in highly ordered tubular crystal arrays rapidly frozen in liquid-nitrogen-cooled ethane. WSA NMR structure was determined at the room temperature in a significantly more dynamic aqueous solution of lyso-lipids; 2% LPPG was needed to maintain monodispersity of the NMR sample when pH was lowered to 5.8. Similarly, GLIC was crystallized at a non-freezing temperature in ~3% n-Undecyl-β-D-maltoside detergent. Moreover, nAChR normally resides in eukaryotic membranes that contain cholesterol, and cholesterol has been proposed to form a slowly exchangeable structural component of this family of ligand-gated ion channels [35]. The absence of cholesterol in WSA may therefore make it more resemblant to the GLIC structure. Hence, the environment might play a more critical role than the amino acid sequences in affecting the high-resolution details of the structures. The debatable question remains as to which biophysical conditions are more relevant for membrane protein structure determination. Before high-resolution structures can be determined under physiologically relevant conditions, all currently available methods, including crystallography, can conceivably introduce some degree of artifact. Second, the high-resolution difference between WSA and cryo-EM nAChR structures may result from different equilibriums of conformational states. The cryo-EM structure is believed to mainly capture the closed state of the channel [3], whereas GLIC crystal is thought to be in either an open or desensitized state [9,10]. Because WSA structure was determined as a monomer, its conformational state can only be inferred. The fact that WSA structure agrees so well with the GLIC structure suggests that WSA monomer may reflect the conformation of the open or desensitized state of nAChR TM domains. Third, there is also a possibility that WSA structure is distorted due to the truncation of extracellular and intracellular domains or due to water solubilization itself. This scenario is very unlikely, however, given the agreement between WSA and GLIC structures. GLIC was determined as an intact channel protein. Taken together, our results suggest that water-solubilization of membrane proteins is not less relevant for structure determination than other more conventional lipid or detergent solubilization approaches. The ease in protein expression and purification and, more importantly, the greatly improved spectral quality of the resulting proteins after water solubilization, make water solubilization an attractive alternative. We believe that the NMR structures of water-solubilized proteins, including WSK3 published previously [13] and WSA determined in this study, represent the first approximation of the native structures. These structures provide the high-resolution templates for further refinement by other biophysical means.

3.4.2 Dynamic Characteristics of WSA

It is conceivable that water solubilization of membrane proteins can change the motional characteristics of these proteins. For WSA, several residues in the TM2 domain were mutated from hydrophobic to charged amino acids for the purpose of promoting monomeric form in order to increase NMR spectral resolution. There are seven mutations in the TM2 helix, four of which, L89K (L9’K), V93E (V13’E), L96K (L16’K), and V97K (V17’K) (see Figure 3.1d for relative numbering of TM2), are located at the putative pore-lining sites of the nAChR channel. The charged residues likely prevent the formation of oligomerization by TM2. Without association with other TM2 helices, TM2 domain becomes particularly dynamic in WSA. Although the increase in dynamics of TM2 in WSA is undoubtedly the result of water solubilization, the fact that other TM domains do not experience the same degree of increase suggests that the motional characteristics of TM2, controlled by Loop-12 and Loop-23, might be intrinsically different from the rest of the protein. Indeed, it has been suggested that nAChR TM domains can be separated into inner helices, comprised of TM2 domains, and outer helices, comprised of TM1, TM3, and TM4 helices [3]. Channel gating is thought to involve the movement of the inner helices relative to the outer helices.

The dynamics of the two loops, Loop-12 and Loop-23, deserves further examination. As shown in Figure 3.1d, the WSA loop regions coincide more with those in GLIC structure than in nAChR cryo-EM structure. Loop-12 consists of five residues L73-S77 (LPTDS) in WSA, different from the very tight, three-residue turn (SGE) in the cryo-EM structure. The loop residues are also shifted. Loop-12 in WSA is a part of the TM1 helix in nAChR cryo-EM structure, whereas Loop-12 in the cryo-EM structure forms the beginning of the TM2 helix in WSA. The NMR dynamics measurements show that the segment from Y72 to K80, encompassing Loop-12 in both the WSA (LPTDS) and cryo-EM (SGE) structures, is highly flexible. Although the Loop-12 conformation as shown in the NMR structures is the preferred conformation based on the NMR data, the shifted Loop-12 conformation as displayed in the Cryo-EM structure is also possible. Nevertheless, we believe that the Loop-12 conformation as shown in the NMR structures, with P74 in the loop instead of in the TM1 helix, is the dominant loop population in the conformational equilibrium.

Similarly, Loop-23 between TM2 and TM3 domain is also highly dynamic. In the intact receptor, this loop is believed to interact with the Cys-loop and β12 loop in the extracellular domain to mediate ligand binding to channel gating [17,18,36,37]. In WSA structure, TM2 helix ends at E20’, whereas in Cryo-EM structure the TM2 helix extends to the extracellular space and ends at S27’ or A28’. None of the high-resolution structures solved so far for Cys-loop receptors, including the WSA structure solved in this study, the NMR structure of the nAChR β2 subunit TM domains [6], the high resolution NMR structures of glycine receptor TM2-TM3 domains solved in DMPC/DHPC bicelles [38], the crystal structure of GLIC [10], and the crystal structure ELIC [8], agree with the cryo-EM structure in this region. Our dynamics data suggest that several conformations can co-exist in Loop-23. Even with water solubilization, the dynamic characteristics seem to have been preserved. High-resolution NMR spectra of pentameric glycine receptor TM domains in lyso-lipid micelles also exhibit peak doubling in this segment (data not shown), suggesting that slow conformational exchange is an intrinsic dynamic feature in the region between TM2 and TM3 helices. The conformation heterogeneity of Loop-23 may be required for the TM2 movement during channel gating.

3.4.3 Anesthetic Binding to WSA

To further confirm similarity of WSA to GLIC, we performed binding studies. Recently, X-ray structures of GLIC were solved with two different general anesthetics—propofol and a haloether, desflurane—bound [39]. Like GLIC, nAChR is also sensitive to general anesthetics, so confirmation of anesthetic binding will provide a functional measure of the fidelity of the water solubilization paradigm. We found that both propofol and a haloether photolabels bound to the same residue (V31) in the WSA structure (Figure 3.4). Also highlighted in this figure are residues that border the propofol-binding site in the GLIC-anesthetic co-crystal structure. The agreement with respect to the Z-direction along the bundle is remarkable. Although the side chain of V31 is oriented away from the intrasubunit GLIC anesthetic site, we have prior evidence suggesting that backbone carbonyl atoms are preferred photoadduction sites [40], especially for residues with apolar hydrocarbon side chains. The V31 carbonyl oxygen is oriented such that it is easily accessed from the intrasubunit cavity. Thus, it would appear that the anesthetic site in GLIC and WSA is similarly positioned within the bundle. This result further validate the water-solubilization approach: mutation of surface residues to improve water solubility does not greatly alter the overall fold and internal packing of the protein, preserving potential drug binding sites or structure-based drug screening. However, it is important to note that neither propofol nor isoflurane binding site has been definitely located in nAChR, thus we cannot yet conclude with confidence that the WSA site well represents that in the native nAChR.

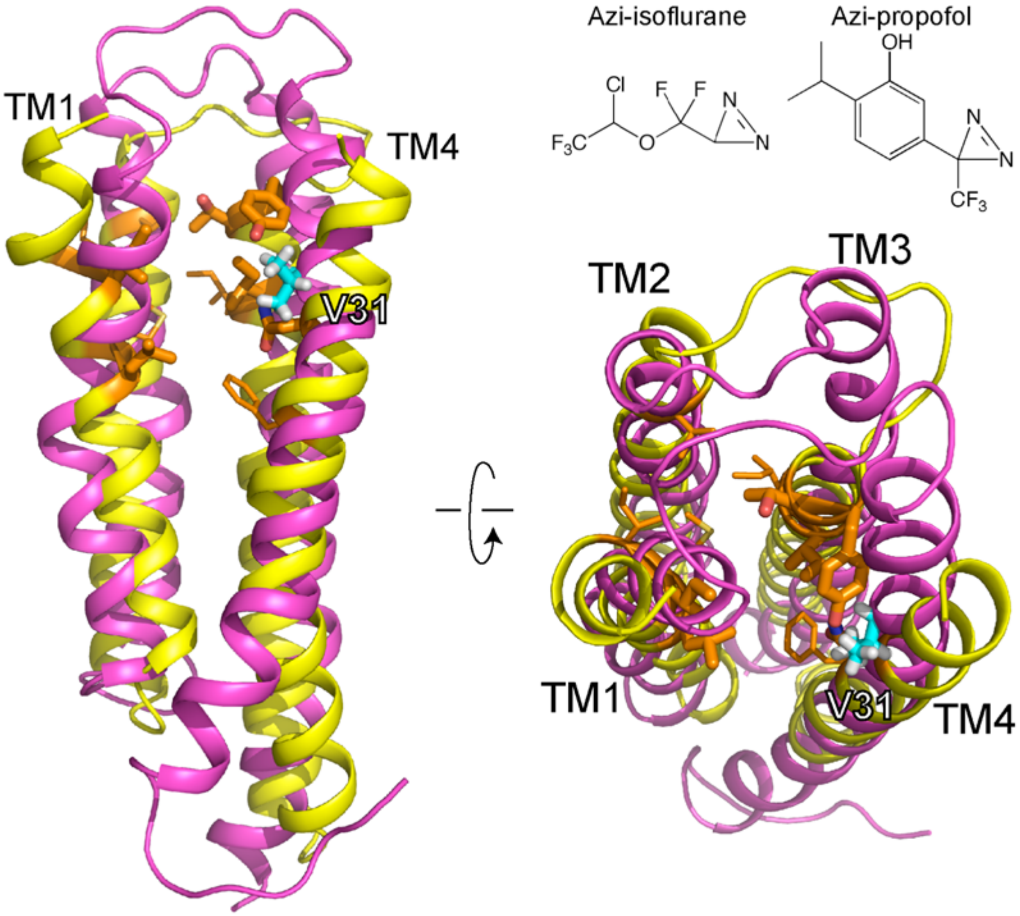


Figure 3.4. Comparison of Anesthetic Binding Site in WSA and in GLIC. The WSA structure (magenta) is fitted to the TM domain of the crystal structure of GLIC (yellow). The anesthetic binding site in WSA was determined by photoaffinity labeling. Both Azi-isoflurane and Azi-propofol bind to residue V31 located in TM4 (cyan). The dominant propofol binding site in GLIC crystal is highlighted in orange. Notice that the general location of the binding sites is essentially the same, suggesting that the drug-binding site is preserved in the water-solubilized protein.

3.4.4 WSA-Detergent Interaction Site

Although WSA was soluble in water at high pH, lowering pH to pass the theoretical isoelectric point (pI) for NMR experiments can lead to soluble aggregations. To maintain WSA monodispersity in the NMR sample, 2% of the lyso-lipid LPPG was used. Previous experimental and computational studies have suggested that nicotinic acetylcholine receptor requires the proper compositions of lipids in order to function. The location of lyso-lipid molecules in WSA thus can provide structural information about protein-lipid interaction. We found that several residues at the TM1 and TM4 interface have cross peaks with LPPG. Specifically, residues 48, 50, 64, and 67 of TM1 and residues 24, 33, 34, 35, 39, 40, 41, and 43 of TM4 border an interfacial cavity that can accept a lipid molecule. The same cavity can be found in nAChR cryo-EM structure and in ELIC and GLIC X-ray structures. As shown in Figure 3.5, the superposition of WSA and GLIC reveals that the strong WSA-LPPG interaction site coincides nearly perfectly with location where surface lipid molecules were found in the GLIC crystal. This finding indicates that despite water solubilization, the helix interfaces are preserved to maintain possible interactions as in the wild-type protein.

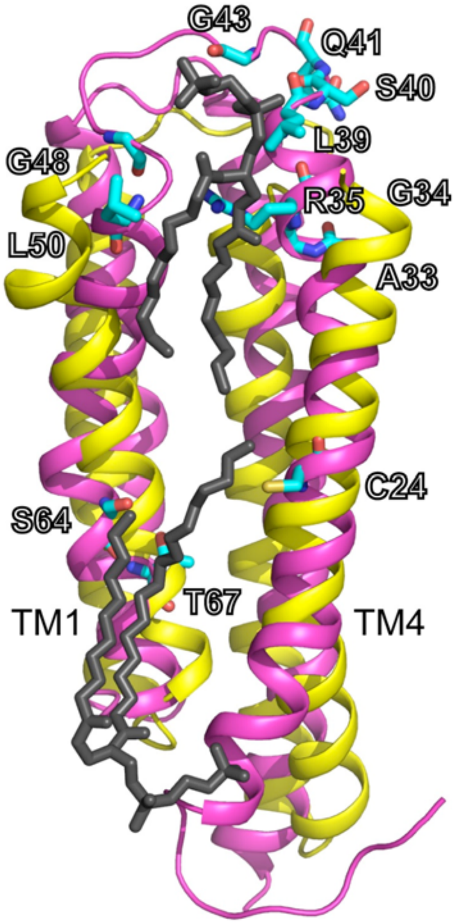


Figure 3.5. WSA-LPPG Interactions. WSA structure is superimposed onto the TM domain of the GLIC crystal structure. Residues showing cross-peaks with LPPG are highlighted in licorice representation in cyan (for clarity no hydrogen atoms are displayed). Also displayed are lipid molecules co-crystallized with GLIC (black). Notice that the intra-helix cavity between TM1 and TM4 for lipid binding is preserved in WSA.

In summary, the monomer structure of a water-soluble analogue of the *Torpedo* nAChR α1 subunit was resolved to the backbone RMSD of 0.69 ± 0.13 Å. The overall structure is similar to the crystal structure of GLIC, demonstrating the robustness of the water solubilization approach. The WSA structure preserved the anticipated intra-helical drug binding site and lipid interaction site in the intact form of homologous receptors.

3.5 References

1. Lindstrom JM: **Nicotinic acetylcholine receptors of muscles and nerves: comparison of their structures, functional roles, and vulnerability to pathology**. *Ann N Y Acad Sci* 2003, **998**:41-52.
2. Baenziger JE, Corringer PJ: **3D structure and allosteric modulation of the transmembrane domain of pentameric ligand-gated ion channels**. *Neuropharmacology* **60**:116-125.
3. Miyazawa A, Fujiyoshi Y, Unwin N: **Structure and gating mechanism of the acetylcholine receptor pore**. *Nature* 2003, **423**:949-955.
4. Unwin N: **Refined structure of the nicotinic acetylcholine receptor at 4A resolution**. *J Mol Biol* 2005, **346**:967-989.
5. Dellisanti CD, Yao Y, Stroud JC, Wang ZZ, Chen L: **Crystal structure of the extracellular domain of nAChR alpha1 bound to alpha-bungarotoxin at 1.94 A resolution**. *Nat Neurosci* 2007, **10**:953-962.
6. Bondarenko V, Tillman T, Xu Y, Tang P: **NMR structure of the transmembrane domain of the n-acetylcholine receptor beta2 subunit**. *Biochim Biophys Acta* **1798**:1608-1614.
7. Brejc K, van Dijk WJ, Klaassen RV, Schuurmans M, van Der Oost J, Smit AB, Sixma TK: **Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors**. *Nature* 2001, **411**:269-276.

8. Hilf RJ, Dutzler R: **X-ray structure of a prokaryotic pentameric ligand-gated ion channel**. *Nature* 2008, **452**:375-379.

9. Bocquet N, Nury H, Baaden M, Le Poupon C, Changeux JP, Delarue M, Corringer PJ: **X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation**. *Nature* 2009, **457**:111-114.

10. Hilf RJ, Dutzler R: **Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel**. *Nature* 2009, **457**:115-118.

11. Slovic AM, Kono H, Lear JD, Saven JG, DeGrado WF: **Computational design of water-soluble analogues of the potassium channel KcsA**. *Proc Natl Acad Sci U S A* 2004, **101**:1828-1833.

12. Bronson J, Lee OS, Saven JG: **Molecular Dynamics Simulation of WSK-3, a Computationally Designed, Water-Soluble Variant of the Integral Membrane Protein KcsA**. *Biophys J* 2006, **90**:1156-1163.

13. Ma D, Tillman TS, Tang P, Meirovitch E, Eckenhoff R, Carnini A, Xu Y: **NMR studies of a channel protein without membranes: structure and dynamics of water-solubilized KcsA**. *Proc Natl Acad Sci U S A* 2008, **105**:16537-16542.

14. Zhou Y, Morais-Cabral JH, Kaufman A, MacKinnon R: **Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 A resolution**. *Nature* 2001, **414**:43-48.

15. Cui T, Canlas CG, Xu Y, Tang P: **Anesthetic effects on the structure and dynamics of the second transmembrane domains of nAChR alpha4beta2**. *Biochim Biophys Acta* **1798**:161-166.

16. Xu Y, Seto T, Tang P, Firestone L: **NMR study of volatile anesthetic binding to nicotinic acetylcholine receptors**. *Biophys J* 2000, **78**:746-751.

17. Liu LT, Haddadian EJ, Willenbring D, Xu Y, Tang P: **Higher susceptibility to halothane modulation in open- than in closed-channel alpha4beta2 nAChR revealed by molecular dynamics simulations**. *J Phys Chem B* **114**:626-632.

18. Liu LT, Willenbring D, Xu Y, Tang P: **General anesthetic binding to neuronal alpha4beta2 nicotinic acetylcholine receptor and its effects on global dynamics**. *J Phys Chem B* 2009, **113**:12581-12589.

19. Chiara DC, Dangott LJ, Eckenhoff RG, Cohen JB: **Identification of nicotinic acetylcholine receptor amino acids photolabeled by the volatile anesthetic halothane**. *Biochemistry* 2003, **42**:13457-13467.

20. Fraczkiewicz R, Braun, W.: **Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules**. *J Comput Chem* 1998, **19**:15.

21. Kono H, Saven JG: **Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure**. *J Mol Biol* 2001, **306**:607-628.

22. Calhoun JR, Kono H, Lahr S, Wang W, DeGrado WF, Saven JG: **Computational design and characterization of a monomeric helical dinuclear metalloprotein**. *J Mol Biol* 2003, **334**:1101-1115.

23. Cochran FV, Wu SP, Wang W, Nanda V, Saven JG, Therien MJ, DeGrado WF: **Computational de novo design and characterization of a four-helix bundle protein that selectively binds a nonbiological cofactor**. *J Am Chem Soc* 2005, **127**:1346-1347.

24. Nanda V, Rosenblatt MM, Osyczka A, Kono H, Getahun Z, Dutton PL, Saven JG, Degrado WF: **De novo design of a redox-active minimal rubredoxin mimic**. *J Am Chem Soc* 2005, **127**:5804-5805.

25. Bender GM, Lehmann A, Zou H, Cheng H, Fry HC, Engel D, Therien MJ, Blasie JK, Roder H, Saven JG, et al.: **De novo design of a single-chain diphenylporphyrin metalloprotein**. *J Am Chem Soc* 2007, **129**:10732-10740.

26. Dunbrack RL, Jr., Cohen FE: **Bayesian statistical analysis of protein side-chain rotamer preferences**. *Protein Sci* 1997, **6**:1661-1681.

27. Weiner SJ, Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P.: **A New Force-Field For Molecular Mechanical Simulation Of Nucleic-Acids And Proteins**. *J. Am. Chem. Soc.* 1984, **106**:20.

28. Fiser A, Do RK, Sali A: **Modeling of loops in protein structures**. *Protein Sci* 2000, **9**:1753-1773.

29. Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S, Hochstrasser D: **The Focusing Positions Of Polypeptides In Immobilized Ph Gradients Can Be Predicted From Their Amino-Acid-Sequences**. *Electrophoresis* 1993, **14**:1023-1031.

30. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K: **Scalable molecular dynamics with NAMD**. *J Comput Chem* 2005, **26**:1781-1802.

31. MacKerell AD, Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M.: **All-atom empirical potential for molecular modeling and dynamics studies of proteins**. *J. Phys. Chem. B* 1998, **102**:31.

32. Gorne-Tschelnokow U, Strecker A, Kaduk C, Naumann D, Hucho F: **The transmembrane domains of the nicotinic acetylcholine receptor contain alpha-helical and beta structures**. *EMBO J* 1994, **13**:338-341.

33. Cascio M, Shenkel S, Grodzicki RL, Sigworth FJ, Fox RO: **Functional reconstitution and characterization of recombinant human alpha 1-glycine receptors**. *J Biol Chem* 2001, **276**:20981-20988.

34. Ma D, Liu Z, Li L, Tang P, Xu Y: **Structure and dynamics of the second and third transmembrane domains of human glycine receptor**. *Biochemistry* 2005, **44**:8790-8800.

35. Brannigan G, Henin J, Law R, Eckenhoff R, Klein ML: **Embedded cholesterol in the nicotinic acetylcholine receptor**. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:14418-14423.

36. Mowrey D, Haddadian EJ, Liu LT, Willenbring D, Xu Y, Tang P: **Unresponsive correlated motion in alpha7 nAChR to halothane binding explains its functional insensitivity to volatile anesthetics**. *J Phys Chem B* **114**:7649-7655.

37. Szarecka A, Xu Y, Tang P: **Dynamics of heteropentameric nicotinic acetylcholine receptor: implications of the gating mechanism**. *Proteins* 2007, **68**:948-960.

38. Canlas CG, Ma D, Tang P, Xu Y: **Residual dipolar coupling measurements of transmembrane proteins using aligned low-q bicelles and high-resolution magic angle spinning NMR spectroscopy**. *J Am Chem Soc* 2008, **130**:13294-13300.

39. Nury H, Van Renterghem C, Weng Y, Tran A, Baaden M, Dufresne V, Changeux JP, Sonner JM, Delarue M, Corringer PJ: **X-ray structures of general anaesthetics bound to a pentameric ligand-gated ion channel**. *Nature* **469**:428-431.

40. Eckenhoff RG, Xi J, Shimaoka M, Bhattacharji A, Covarrubias M, Dailey WP: **Azi-isoflurane, a Photolabel Analog of the Commonly Used Inhaled General Anesthetic Isoflurane**. *Acs Chemical Neuroscience* 2010, **1**:139-145.

4 Comparative Modeling of the Human μ Opioid Receptor: A Structural Context for Mutagenesis Studies and Agonist-Induced Activation[1]

4.1 Introduction

The study of opioids, their effects, and their interactions with opioid receptors [1,2] are relevant to addiction, pain control and reward pathways [1,3,4]. Clinically, opioid receptors are important in analgesia, euphoria, respiratory depression, feeding, hormone release, inhibition of gastrointestinal transit, and anxiety [5,6]. To better understand the pharmacological activity and diversity of the opioid receptors, it is desirable to have molecularly detailed structures that aid in understanding their cellular responses to a variety of ligands. Such structural information could also aid the rational design of new drugs. Until now, the structures of the opioid receptors have not been obtained by experimental means, and therefore computational models are of central importance. Herein comparative modeling is used to obtain models of the μ opioid receptor and better understand some of its functionally related properties.

---

[1] This project was a collaboration with the group of Professor Renyu Liu at the Department of Anesthesiology and Critical Care, University of Pennsylvania School of Medicine, Philadelphia, PA 19104.

Opioid receptors are part of the largest family of integral transmembrane proteins coded by the human genome, the G protein-coupled receptors (GPCRs) [7]. GPCRs mediate most transmembrane signal transduction, usually in response to hormones, neurotransmitters and environmental stimulants. Each GPCR comprises an extracellular N terminus, seven-transmembrane (7TM) helical segments separated by alternating intracellular and extracellular loop regions, and an intracellular C terminus [7-9]. Opioid receptors are part of the largest GPCR family, *family A* (rhodopsin-like) [10], and other members include the receptors for epinephrine, dopamine, serotonin, and adenosine [11].

The μ opioid receptor is the primary receptor in the brain for endogenous opioid neuropeptides as well as exogenously administrated opioid compounds [1,3,12]. Potent drugs such as morphine, heroin, fentanyl and methadone induce their pharmacological effects through the activation of this receptor [13].

The μ opioid receptor structure has not yet been experimentally determined, and indirect computational methods, including comparative modeling and de novo methods, have been used to suggest structural details of this important signal transduction protein [14-19]. Comparative (homology) modeling provides three-dimensional information concerning protein structures, particularly for proteins such as the μ opioid receptor that are difficult to obtain experimentally in large quantity for use in X-ray crystallography or protein NMR structural studies [20,21]. The atomic-resolution structural information suggested from such models can be used to interpret experimental data, predict the functional effects of different mutations, motivate hypotheses concerning structure and function, and design medications [22].

In creating models of the opioid receptors, only the structure of bovine rhodopsin has been used previously [15-17]. In recent years, several additional GPCR structures have been solved to high resolution: the human β_2 adrenergic receptor (2.4 Å) [23], the turkey β_1 adrenergic receptor (2.7 Å) [24], and the human A_{2A} adenosine receptor (2.6 Å) [25]. The use of these additional GPCR structures may improve the quality of the models and be useful in formulating hypotheses and interpreting mutational data. The use of the structure of the β_2 adrenergic receptor seems particularly appealing [23,26]. Although the sequence identity between the human μ opioid receptor and the human β_2 adrenergic receptor is just (26%), compared to that with bovine rhodopsin (25%), the μ opioid receptor and β_2 adrenergic receptor may share similar interactions with their respective ligands since epinephrine is structurally similar to opioids [1,27]. In addition, increasing the number of templates often helps to improve the accuracy of comparative protein modeling [28,29]. Comparative models created by using just one template emphasize the shortcomings of the modeling procedure by heavily biasing the structures of the created models [23].

Herein models of the human μ opioid receptor (hMOP-R) were constructed with comparative modeling. Two families of structures were created: (a) one used structures of human β_2 adrenergic receptor and bovine rhodopsin as templates and (b) the other also included the turkey β_1 adrenergic receptor and the human A_{2A} adenosine receptor as templates. Representative structures of each of these two- and four-template ensembles were considered in light of a large set of mutagenesis studies. The two- and four-template models are very similar and may be used to interpret a variety of experimental data. Molecular docking calculations were carried out on these models using an opioid antagonist (naloxone) and an opioid agonist (morphine). The docking calculations are consistent with previously proposed binding sites and suggest specific interactions of the opioid ligands with key residues in the receptor. Using the proposed agonist-induced activation mechanisms of the adrenergic receptors and the A_{2A} adenosine receptor, residues involved in agonist recognition of h-MOR-P are suggested.

4.2 Methods

The approach taken to obtain models of hMOP-R involves alignment of the sequences, creation of the comparative models, minimization, and structural evaluation. Docking calculations were used to assess the models and potential protein-ligand interactions.

4.2.1 Sequence Alignment

The sequences of hMOP-R, human β_2 adrenergic receptor, bovine rhodopsin, turkey β_1 adrenergic receptor and human A_{2A} adenosine receptor were obtained from the UniProt Knowledgebase UniProtKB server with the accession numbers P35372, P07550, P02699, P07700, and P29274, respectively [30].

Given the importance of sequence alignments in the comparative modeling procedure [20,31,32], several different programs and substitution matrices were considered. *BLASTp* [33], *SIM* [34], *ClustalW* [35], and *Phyre* [36] were used to align the sequences of hMOP-R and human β_2 adrenergic receptor. In the case of *BLASTp*, three members of the “Blosum” substitution matrix family [37] (Blosum62, Blosum45 and Blosum80) and one member of the “PAM” substitution matrix family [38] (PAM70) were used. For the case of the alignment tool *SIM*, Blosum62 and Blosum30 were used. For *ClustalW* just the Blosum30 matrix was used. The protein structure prediction server *Phyre*, was also utilized. Standard penalty gaps were applied in all the cases [33-35]. A similar procedure was carried out in the case of hMOP-R and bovine rhodopsin, hMOP-R and β_1 adrenergic receptor, and hMOP-R and A_{2A} adenosine receptor.

A final sequence alignment was obtained for each pair of proteins and modifications were performed to maintain highly conserved fingerprint residues of the rhodopsin-like GPCR family [39]. Among these are: the *disulfide bond* between TM3 and the second extracellular loop (EC2), the “*DRY”* motif in TM3, the *XBBXXB* motif in the third intracellular loop (IC3) (where *B* represents a basic amino acid and *X* represents a non-basic residue, *LRRITR* in the case of hMOP-R), the *FXXXWXPX[F]* motif in TM6 (*FIVCWTPIH* in the case of hMOP-R), the *NPXXY* motif in TM7 (*NPVLY* in hMOP-R), and the C-terminal cys palmitoylation site.[39] The final multiple sequence alignment is presented in Figure 4.1.


```
          70        80        90        100       110       120       130
hMOP-R   SMITAITIMALYSIVCVVGLFGNFLVMYVIVRYTKMKTATNIYIFNLALADALATS-TLPFQSVNYLMGTWPFG
ADRβ2    DEVWVVGMGIVMSLIVLAIVFGNVLVITAIAKFERLQTVTNYFITSLACADLVMGLAVVPFGAAHILMKMWTFG
bRHO     EPWQFSMLAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFG
ADRβ1    --QWEAGMSLLMALVVLLIVAGNVLVIAAIGRTQRLQTLTNLFITSLACADLVMGLLVVPFGATLVVRGTWLWG
AA2AR    -IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAITISTGFCAACH

         140       150       160       170       180       190       200       210
hMOP-R   TILCKIVISIDYYNMFTSIFTLCTMSVDRYIAVCHPVKALDFRTPRNAKIINVCNWILSSAIG-LPVMFMATTK
ADRβ2    NFWCEFWTSIDVLCVTASIETLCVIAVDRYFAITSPFKYQSLLTKNKARVIILMVWIVSGLTSFLPIQMHWYRA
bRHO     PTGCNLEGFFATLGGEIALWSLVVLAIERYVVVCKPMS-NFRFGENHAIMGVAFTWVMALACA-APPL-VGWSR
ADRβ1    SFGCECWTSLDVLCVTASIETLCVIAIDRYLAITSPFRYQSLMTRARAKVIICTVWAISALVSFLPIMMHWWRD
AA2AR    --GCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWVLSFAIGLTPMLGWNNCG

                     220       230          240       250       260        270
hMOP-R   Y-----------RQGSIDCTLTFSHPTWYW---ENLLKICVFIFAFIMPVLIITVCYGLMILRLK-SVRMLSGS
ADRβ2    T-----------HQEAINCYANETCCDFFT---NQAYAIASSIVSFYVPLVIMVFVYSRVFQEAK-RQLQKIDK
bRHO     Y-----------IPEGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFFCYGQLVFTVK--EAAAQQQ
ADRβ1    E-----------DPQALKCYQDPGCCDFVT---NRAYAIASSIISFYIPLLIMIFVYLRVYREAK-EQIRKIDR
AA2AR    QPKEGKNHSQGCGEGQVACLFEDVVPM-------NYMVYFNFFACVLVPLLLMLGVYLRIFLAAR-RQLKQMES

                                                   280       290       300
hMOP-R   K-----------------------------------EKDRNLRRITRMVLVVVAVFIVCWTPIHIYVIIKALVT
ADRβ2    SEGRFHVQNLSQVEQDGRTGHGLRRSSKFC---------LKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQD
bRHO     ESA---------------------------------TTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQ
ADRβ1    CEGRFYGSQEQPQPPPLPQHQPILGNGRASKRKTSRVMAMREHKALKTLGIIMGVFTLCWLPFFLVNIVNVFNR
AA2AR    QPLPGERAR---------------------------STLQKEVHAAKSLAIIVGLFALCWLPLHIINCFTFFCP

        310       320       330       340       350
hMOP-R   IPETTFQTVSWHFCIALGYTNSCLNPVLYAFLDENFKRCFRE-FC
ADRβ2    NL---IRKEVYILLNWIGYVNSGFNPLIYC-RSPDFRIAFQELLCL
bRHO     GSD--FGPIFMTIPAFFAKTSAVYNPVIYIMMNKQFRNCMVTTLCC
ADRβ1    DL---VPDWLFVFFNWLGYANSAFNPIIYC-RSPDFRKAFKRLLCF
AA2AR    DCSH-APLWLMYLAIVLSHTNSVVNPFIYAYRIREFRQTFRK-IIRSHVLRQ
```

Figure 4.1. **Sequence Alignment Used in the Creation of the Models of the Human μ Opioid Receptor, hMOP-R.** The templates are: human β_2 adrenergic receptor (ADRβ2) and bovine rhodopsin (bRHO). The residues of the N- and C- terminus are excluded (residue 1 to 65 and residues 354 to 400, respectively). Also, the residues excluded from the comparative modeling are colored in gray. The most conserved residues at each of the transmembrane helices are depicted in blue. The secondary structure of the β_2 adrenergic receptor based on STRIDE [40], is shown below the sequences. Residue numbering of hMOP-R is shown. Highly conserved motifs in the rhodopsin-like GPCR family are highlighted in yellow.

4.2.2 Comparative Modeling and Minimization

Two different ensembles of models of the hMOP-R were constructed. (i) The first ensemble used the X-ray crystallographic structures of human β_2 adrenergic receptor at 2.4 Å resolution (PDB accession code: 2RH1) [23] and bovine rhodopsin at 2.2 Å resolution (PDB accession code: 1U19) [41] as templates. (ii) In addition to these templates, the second ensemble used the X-ray crystallographic structures of turkey β_1 adrenergic receptor at 2.7 Å resolution (PDB accession code: 2VT4) [24] and human A_{2A} adenosine receptor at 2.6 Å resolution (PDB accession code: 3EML) [25].

Using the sequence alignments and the two- and four-template structure sets described above, one hundred independent models (from residue 65 to 353) were generated using Modeller 9v2 with the refinement optimization level adjusted to *slow* [42,43]. The side chains from the resulting models of the two-template and four-template ensembles, were minimized using NAMD2 [44] and the CHARMM22 force field [45]. Hydrogen atoms were added and minimization was performed by the conjugate-gradient method until the total energy remains constant (change in energy less than 1.0 kcal/mol). Models with the lowest energy were characterized using Molprobity [46] to confirm that no steric clashes or unusual conformations of the backbone and side chains are present. The binding site cavities for the models were calculated and analyzed with CASTp [47].

4.2.3 Docking Calculations

Naloxone (antagonist) and morphine (agonist) were docked into representative models using DockingServer [48] (the DockingServer uses the AutoDock4 [49] algorithm for the docking calculations). The geometries of naloxone and morphine were obtained from the PubChem database (CID: 5284596 and 5288826, respectively) [50]. The geometry of the ligands was optimized using the PM3 semiempirical method using MOPAC2009 [51], which was also used to calculate the partial charges. The Lamarkian genetic algorithm was used during the docking calculations with the local search method developed by Solis and Wets [52]. The initial position, orientation, and value of the torsion angles for the ligands were set randomly. Default parameters for the genetic algorithm were used in all the calculations.

4.3 Results and Discussion

4.3.1 Template Structures

The four structural templates used for the creation of the hMOP-R belong to the same family of hMOP-R (*family A*) and are among the most studied GPCR systems. Each receptor comprises seven transmembrane helices that form a helical bundle. The helices are connected by intracellular (IC) and extracellular (EC) loops.

The human β_2 adrenergic receptor is part of the *family A* of GPCRs, particularly of the amine group [23,26,53], and is the target of the catecholamine hormone epinephrine (adrenaline). The structure contains a small helical segment perpendicular to the seven-helix bundle. Helices TM2, TM4, TM5, TM6 and TM7 contain proline-induced kinks that are believed to be relevant in the rearrangement needed for the activation of the G protein. An α-aneurysm is present in helix TM2, and the structure has an exposed loop (EC2) that forms a helical segment.

Rhodopsin is composed of the protein opsin and the covalently bound cofactor retinal [41]. Proline-induced kinks are located in helices TM1, TM4, TM6 and TM7. In contrast to the second extracellular loop in β_2 adrenergic receptor, the EC2 loop forms a partially buried β-sheet structure. Similar to the β_2 adrenergic receptor, there is an α-aneurysm present in TM2.

The turkey β_1 adrenergic receptor shares high similarity with the β_2 adrenergic receptor [24]. Structurally, one of the main differences is a helical motif located in the second intracellular loop (IC2) that is absent in the structure of β_2 adrenergic receptor and also absent in the structure of bovine rhodopsin.

The human A_{2A} adenosine receptor endogenous ligand is the nucleoside adenosine [25]. Proline-induced kinks are located in helices TM2, TM5, TM6 and TM7 and in addition to the highly conserved disulfide bond, two extra disulfide bonds restrict the conformational space of the EC2 loop. Similar to the β_1 adrenergic receptor, a helical segment located in the second intracellular loop (IC2) that is absent in the β_2 adrenergic receptor and bovine rhodopsin. The protein also contains an α-aneurysm in TM2.

These four structures were used as templates in the comparative modeling procedure. Two different sets of models of hMOP-R were created in order to test the hypothesis that more templates produce better models. The first set contains an ensemble of structures obtained using two templates, the β_2 adrenergic receptor and bovine rhodopsin. The second set contains an ensemble of structures generated using four templates: β_2 adrenergic receptor, bovine rhodopsin, β_1 adrenergic receptor and A_{2A} adenosine receptor. Importantly, in order to isolate if significant differences exist with the introduction of more templates, the sequence alignments used were exactly the same in both calculations. The conserved disulfide bond was explicitly specified in the calculation. Special attention was paid to relax the α-aneurysm in TM2 present in all the templates and suggested to be absent in the μ opioid receptor [17]. Finally, the structures were refined by energy minimization (see Methods).

4.3.2 μ Opiod Receptor Mutagenesis Studies

Site-directed mutagenesis studies provide valuable molecular information and can be compared with the comparative models. Although the human μ opioid receptor has been widely studied, the rat and mouse μ opioid receptors are the more commonly used to probe the impact of specific mutations. The sequence homology between human, rat and mouse μ opioid receptors is high. For the human and rat μ receptors, the sequence similarity is 94% (376/400) for the entire sequence and 98% (285/288) for the transmembrane portion considered in the present study based on the pairwise alignment using *BLASTp* [33]. The similarity of the human and mouse μ receptors is 94% (376/400) for the entire sequence and 98% (284/288) for the transmembrane domain. Table 4.1 summarizes previous findings from the mutational studies of the μ opiod receptor, and Figures 4.2 and 4.3 illustrate some of the residues involved in the mutagenesis studies as well as highly conserved residues in hMOP-R.

4.3.3 Ensemble Structures

Figures 4.2a and 4.2b render the backbone root mean square deviation (rmsd) for each residue. The calculation is done relative to the lowest energy structure and the average is over models in each ensemble. In the case of four-template ensemble the range of rmsd values is 0.12 to 7.07 Å with an average residue rmsd of 0.44 ± 0.68 Å while in the case of two-template ensemble the range of values is 0.11 to 9.01 Å with an average residue rmsd of 0.53 ± 0.95 Å. As expected, the variability of the transmembrane region is small when compared with that of the loops and the termini. Figures 4.2c and 4.2d show views of relevant ligand-binding residues from the extracellular side of the four-template structure and the two-template structures, respectively. The average side chain rmsd value of all the residues in Table 4.1 is 0.97 ± 0.64 Å for the four-template ensemble and 1.10 ± 0.55 Å for the two-template ensemble. In summary, including two additional templates (β_1 adrenergic receptor and A_{2A} adenosine receptor) does not significantly change the ensemble results when the same alignment is used.

SEQUENCE NUMBER IN THE HMOP-R	DESCRIPTION
D116 (2.50)	Mutation to A, N or Q reduced binding of DAMGO to less than 5% of WT. Also Mutation to N reduces dramatically the affinities for a peptide agonist [54].
D149 (3.32)	Site-directed mutagenesis studies indicated that it is involved in the primary binding site and also that mutation to A and N diminishes the binding affinity to agonist and antagonist. Mutation to E results in similar properties as the wildtype [54,55].
Y150 (3.33)	Mutation to F reduced the binding affinities to some μ selective agonists but did not alter the affinities to DAMGO, naloxone and the non-selective opioids etorphine and buprenorphine [56].
N152 (3.35)	Mutation to A produced a 20-fold increase in binding affinity for fentanyl but just a 10-fold increase for morphine [57].
D166 (3.49)	Mutation to H, Q, Y and M led to agonist independent activation [58,59].
S198 (4.54)	Important not for ligand binding but most likely for maintaining the receptor conformation [60].
I200 (4.56)	Mutation to V resulted in a four- to five-fold decrease affinity for μ agonists [57].
N232 (5.36)	Mutations to T or L increased the potency of fentanyl and morphine [61].
K235 (5.39)	Covalently labeled by β-FNA, as suggested by site-directed mutagenesis studies [62].
T281 (6.34)	Mutation to K at the junction of IC3 loop led to agonist activation [63].
H299 (6.52)	Mutation to A decreases the ligand binding of μ and δ opioid receptors. Mutation to A reduces binding to DAMGO and naloxone [64].
K305 (6.58)	Mutation to E (to mimic κ receptor) produces activity similar to that of wildtype κ receptor for nor-BNI and GNTI (κ receptor selective ligands) [65].
W320 (7.35)	Mutation to A increases the binding affinities for the antagonists NTI and SIOM. Also, mutation to L confers δ receptor-like potency for morphine [66,67].
H321 (7.36)	Mutation to A reduces the binding affinities of opioid ligands but did not alter the affinities of naloxone and bremazocine [56].
I324 (7.39)	In κ receptor, mutation to T produces a significant reduction of the binding of dynorphin-related peptides, indicating that this site is part of a key opioid peptide binding pocket [68].
Y328 (7.43)	Mutation to F decreases the binding affinity of naltrexone. Also this mutation decreases the K_i value of DAMGO by 20-fold [57].
S331 (7.46)	Mutation to A decreases the potency of DAMGO. Together with other findings, it suggests hydrophilic interactions between S331 and DAMGO [69].
C142 (3.25) C219 (EC2)	Mutation of these cysteine residues suggests that the disulfide bond formed is critical for specific ligand binding [70].
H225 (EC2)	This position is critical for the receptor binding. Mutation to S decreases the affinity of bremazocine 25-fold [71].
E231 (EC2)	Mutation to D increases binding affinity of an analogue of the peptide agonist JOM6 [72].
N88 (1.50) D116 (2.50) N334 (7.49)	Mutagenesis studies suggested these highly conserved residues might be part of a hydrogen-bonding network that ties together helices TM1, TM2 and TM3 [15,73].
K235 (5.39) C237 (5.41)	These positions are the cross-linking sites for the reporter affinity label NNA, as suggested by mutagenesis studies. [74].
I236 (5.40) F239 (5.43) H299 (6.52) V302 (6.55)	Based on engineered metal-binding sites (Zn^{2+}) it is suggested that I236 and F239 face the binding-site crevice and form a metal-binding center with H299 and V302 [17].
F243 (5.47) W295 (6.48)	Mutation experiments suggested that these residues might help to form the putative hydrophobic pocket and constitute a general binding domain for opioids based of similarities with the δ opioid receptor [75].
K305 (6.58) V318 (7.33) W320 (7.35) H321 (7.36)	Mutational studies showed the importance of these third extracellular loop residues to differentiate μ from κ opioid receptors [76].
N334 (7.49) D116 (2.50)	Mutation N334D eliminated detectable binding of DAMGO, but an additional mutation of D116N could partially restore the affinity, indicating a structural relation between these two residues [73].

Table 4.1. μ **Opioid Receptor Mutagenesis Findings**. The residue numbers used in the table are those of hMOP-R. The numbering in parentheses is consistent with the Ballesteros and Weinstein indexing system [77]: (number of the transmembrane helix).(residue number relative to most conserved residue in transmembrane helix, which is assigned position 50). EC denotes extracellular loop; IC denotes intracellular loop. β-FNA and NNA are irreversible μ opioid receptor ligands (reporter affinity labels). DAMGO ([D-Ala^2,MePhe^4,Gly-ol^5]enkephalin) is a synthetic opioid peptide having high specificity for μ opioid receptor. NTI is the delta antagonist naltrindole, and SIOM is the delta antagonist 7'-spiroindanooxymorphone. JOM6 (Tyr^1-c(S-Et-S)[D-Cys^2-Phe^3-D-Pen^4]NH_2) is a cyclic peptide agonist.

4.3.4 Model Structures of hMOP-R

The models with the lowest energy based on the CHARMM22 force field were selected as representative of the four-template (4T-hMOP-R) and the two-template (2T-hMOP-R) ensembles. These models of the hMOP-R consist of 288 amino acid residues (residues 66 to 353) comprising the seven transmembrane helices with all the connecting loops, plus the small helix that comes after the seventh transmembrane helix (residues 343-352). N-terminal (residues 1-65) and C-terminal (residues 353-400) domains are not included in the models. The structure of hMOP-R presents the typical counterclockwise arrangement of the seven transmembrane helices (see Figures 4.2c, 4.2d, 4.3b and 4.3c). Proline-induced kinks are located in helices TM2, TM4, TM5, TM6 and TM7. The EC2 loop forms a small helical segment that is inherited from the structures of the adrenergic receptors and is believed to be important in binding selectivity [1]. This loop is part of the entrance to the putative opioid binding pocket and is constrained by a conserved disulfide bond between residues C142 and C219. This extracellular segment is more extended when compared with the comparable tightly packed loop of rhodopsin. The models present a helical segment in IC1 that is present in the all the templates but absent in the structure of bovine rhodopsin. The TM2 in the hMOP-R models does not have the α-aneurysm present in the templates [17].

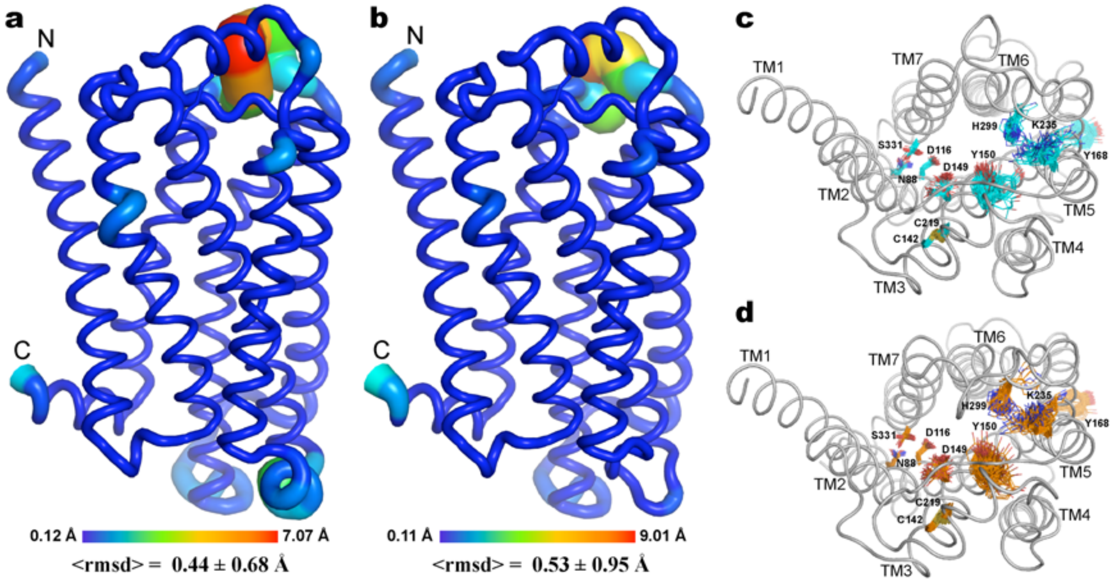


Figure 4.2. Ensemble Structures of hMOP-R Using Four (a) and Two (b) Templates. The thickness of the protein backbone chain and color is representative of the backbone rmsd value for each residue. Subset of some of the residues from Table 4.1 for the 100 ensemble minimized structures for the four-template (**c**) and two-template (**d**) ensemble, respectively. The average side chain rmsd of all the residues in Table 1 is 0.97 ± 0.64 Å for the four-template ensemble and 1.10 ± 0.55 Å for the two-template ensemble.

Two different views of the model of hMOP-R are depicted in Figure 4.3a. In general, 4T-hMOP-R and 2T-hMOP-R are very similar with a backbone rmsd of 1.30 Å. One of the most significant differences between the models is a helical segment in IC2 that is specific to the β_1 adrenergic receptor and A_{2A} adenosine receptor; it is absent in 2T-hMOP-R (see inset of Figure 4.3a). This loop connects helices TM3 and TM4 and is close to the important and highly conserved *DRY* motif.

The secondary structures of the models were assigned with STRIDE [40], a structurally base secondary structure method. The locations and lengths of the TM helices were similar for both 4T-hMOP-R and 2T-hMOP-R.

Both models 4T-hMOP-R and 2T-hMOP-R can qualitatively aid the interpretation of the experimental data collected in Table 4.1 (see Figures 4.3b and 4.3c). Asp149$^{3.32}$, known to be crucial for binding affinity of agonists and antagonists [54,55], is located in TM3 and it is directed toward the putative ligand-binding pocket. Lys235$^{5.39}$ is predicted to appear in this binding site, where it could irreversibly bind the fumaramate moiety of the reporter affinity label β-funaltrexamine (β-FNA) [62]. The models contain a hydrogen-bonding network formed by Asn88$^{1.50}$, Asp116$^{2.50}$ and Asn334$^{7.49}$, and interactions among these residues have been suggested to tie TM1, TM2 and TM3 together [15,73]. Phe243$^{5.47}$ and Trp295$^{6.48}$ have been suggested to form a hydrophobic pocket [75], observed in the models where these residues appear near the interface between helices TM5 and TM6. The residue His299$^{6.52}$ is positioned near both these residues in the putative binding site and is known to be important in agonist affinity [64]. Position His225, in the EC2 loop, is located at the entrance of the binding pocket, and this residue may interact with ligands; this position is critical for the receptor binding [71]. Mutation of Asp166$^{3.49}$ or Thr281$^{6.34}$ leads to an agonist-independent activation of the receptor [58,63]. Residue Asp166$^{3.49}$ is part of the conserved triplet of residues, the *DRY* motif. In the case of 4T-hMOP-R, Asp166$^{3.49}$ is in proximity to Arg181(IC2) (distance of 2.7 Å), Tyr108$^{2.42}$ (4.1 Å), and Thr105$^{2.39}$ (5.5 Å) (see Figure 4.3d). In the case of 2T-hMOP-R, Asp166$^{3.49}$ is near to Arg167$^{3.50}$ (2.7 Å), and is part of a polar network together with Arg181(IC2), Asp179(IC2), and to a less extent with Tyr108$^{2.42}$ (see Figure 4.3e). (The distances in parentheses are the shortest distances between atoms in the different residue side chains.)

The importance of residue Thr281$^{6.34}$ in hMOP-R seems to be related with the region denominated "ionic lock" [78]: the highly conserved Arg167$^{3.50}$ on TM3 together with a Glu$^{6.30}$ residue on TM6 has been proposed to be relevant to the stability of the inactive-state conformation and therefore with the basal activity of different GPCRs [7,79]. In the case of the X-ray crystallographic structure of the inactive state of rhodopsin [41], Arg$^{3.50}$ forms a salt bridge with Glu$^{6.30}$. Surprisingly, in the case of both adrenergic receptors [23,24] and the adenosine receptor [25], the salt bridge between Arg$^{3.50}$ and Glu$^{6.30}$ is disrupted. hMOP-R contains the highly conserved Arg$^{3.50}$ on TM3 but it contains a Leu$^{6.30}$ residue in the analogous position on TM6. In the case of 4T-hMOP-R, Arg167$^{3.50}$ is predicted to be in the vicinity of residue Thr281$^{6.34}$ (2.9 Å), that is, one turn in the helix separated from Leu277$^{6.30}$ (*i*+4). Arg167$^{3.50}$ is also close to the carbonyl group of Leu277$^{6.30}$ (3.0 Å), and to Thr105$^{2.39}$ (4.9 Å). The position of Arg167$^{3.50}$ in 4T-hMOP-R resembles more the position of the equivalent Arg$^{3.50}$ in rhodopsin. In the case of 2T-hMOP-R, the position of Arg167$^{3.50}$ is closer to Asp166$^{3.49}$ (2.7 Å), Thr105$^{2.39}$ (3.6 Å), and the carbonyl group of Ala177(IC2) (4.1 Å) than to Thr281$^{6.34}$ (5.9 Å) as seen in the case of 4T-hMOP-R. The position of Arg167$^{3.50}$ in 2T-hMOP-R resembles the analogous Arg$^{3.50}$ in the adrenergic receptors and the adenosine receptor structures.

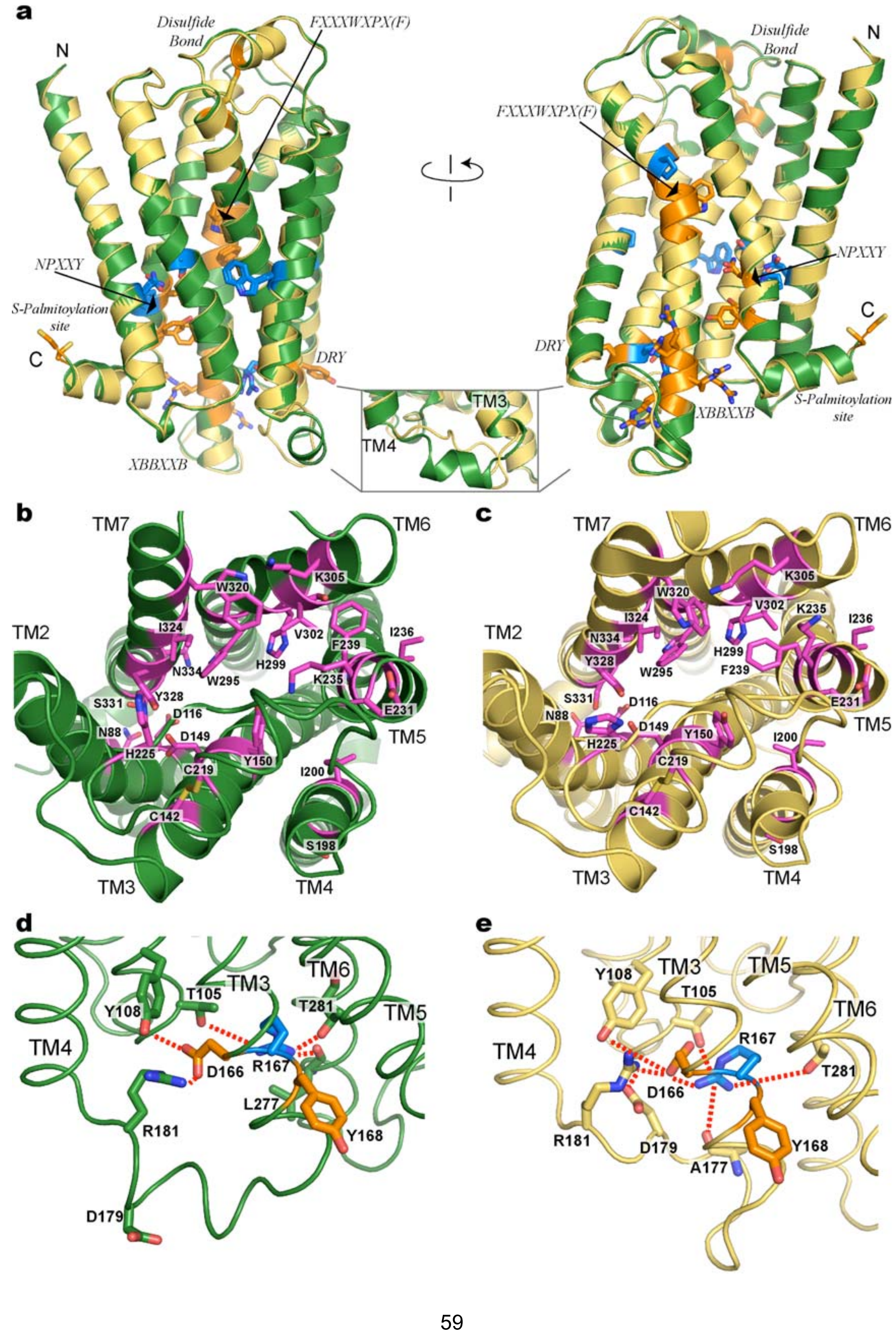

a
Disulfide Bond
FXXXWXPX(F)
N
NPXXY
S-Palmitoylation site
C
DRY
XBBXXB
TM3
TM4
FXXXWXPX(F)
Disulfide Bond
N
NPXXY
C
DRY
XBBXXB
S-Palmitoylation site
b
TM7
TM6
TM2
K305
W320
V302
F239
I236
I324
N334
H299
W295
K235
Y328
S331
D116
E231
N88
D149
TM5
H225
C219
Y150
I200
C142
S198
TM3
TM4
c
TM7
TM6
TM2
K305
W320
V302
K235
I236
I324
N334
H299
Y328
W295
F239
S331
D116
N88
D149
Y150
E231
TM5
H225
I200
C219
C142
S198
TM3
TM4
d
TM6
Y108
T105
TM3
T281
TM5
TM4
D166
R167
L277
R181
Y168
D179
e
Y108
TM3
TM5
T105
TM6
R167
TM4
T281
D166
R181
D179
Y168
A177

Figure 4.3. Alignment and Comparison of Representative Models of hMOP-R. (**a**) the representative models from the four-template ensemble (4T-hMOP-R) and two-template ensemble (2T-hMOP-R) are color in green and yellow, respectively. The most conserved residues at each of the transmembrane helices are depicted as blue sticks. The highly conserved motifs in the rhodopsin-like GPCR family are depicted as orange sticks; disulfide bond between TM3 and EC2, DRY in TM3, XBBXXB in IC3 (where B represents a basic amino acid and X represents a non-basic residue), FXXXWXPX[F] in TM6, the NPXXY in TM7 and the C-terminal cys palmitoylation site. The inset at the bottom shows one of the most significant differences between the two models. The IC2 loop that connects helices TM3 and TM4 in the 4T-hMOP-R model, forms a helical motif inherit from the β_1 adrenergic receptor and the A_{2A} adenosine receptor templates. (**b**) (**c**) Top views of the binding site for the 4T-hMOP-R and 2T-hMOP-R models with a subset of the relevant residues involved in experimental studies (see Table 4.1). Side chain residues are colored in magenta in both cases. This view of the hMOP-R from the extracellular side of the membrane shows the counterclockwise arrangement of the TM helices. (**d**) (**e**), Views of the TM3-IC2-TM4 region with the DRY motif for 4T-hMOP-R and 2T-hMOP-R, respectively. Asp166$^{3.49}$ (D) and Tyr168$^{3.51}$ (Y) are colored in orange while Arg167$^{3.50}$ (R) is colored in blue.

A graphical summary (snake-like plot) of the features of 4T-hMOP-R is shown in Figure 4.4. The lipid membrane region for the h-MOP-R is also shown and is identified using the positioning of the β_2 adrenergic receptor presented in the Orientation of Proteins in Membrane Database [80]. Highly conserved positions, conserved motifs in the rhodopsin-like family, positions involved in experimental studies, length of each of the helices and complete sequence of hMOP-R, are also depicted.

Within the models, CASTp identified a putative binding cavity using a probe with a radius of 1.4 Å. The calculated volume for 4T-hMOP-R is 1402 Å^3 and for 2T-hMOP-R is 1717 Å^3. By comparison, the respective cavity volumes of β_2 adrenergic receptor and bovine rhodopsin are 2057.3 and 674.4 Å^3.

The quality of the models was examined with Molprobity. The Ramachandran analysis for 4T-hMOP-R found one outlier, Thr313 located at the EC3 loop. The clashscore for all-atom contacts was 6.35 (within the 89th percentile). In the case of 2T-hMOP-R, the Ramachandran analysis identified three outliers, Pro226 in EC2, Met226 in IC3, and Thr317$^{7.32}$ at the beginning of TM7. The clashscore for all-atom contacts was 10.17 (within the 70th percentile). As a reference, the clashscores for all-atoms for the templates are: 6.10 (90th percentile) for β_2 adrenergic, 43.86 (6th percentile) for bovine rhodopsin, 19.05 (35th percentile) for β_1 adrenergic and 11.17 (66th percentile) for A_{2A} adenosine receptors. The number of outliers in the Ramachandran analysis for the four templates is zero, eight, one and zero, respectively.

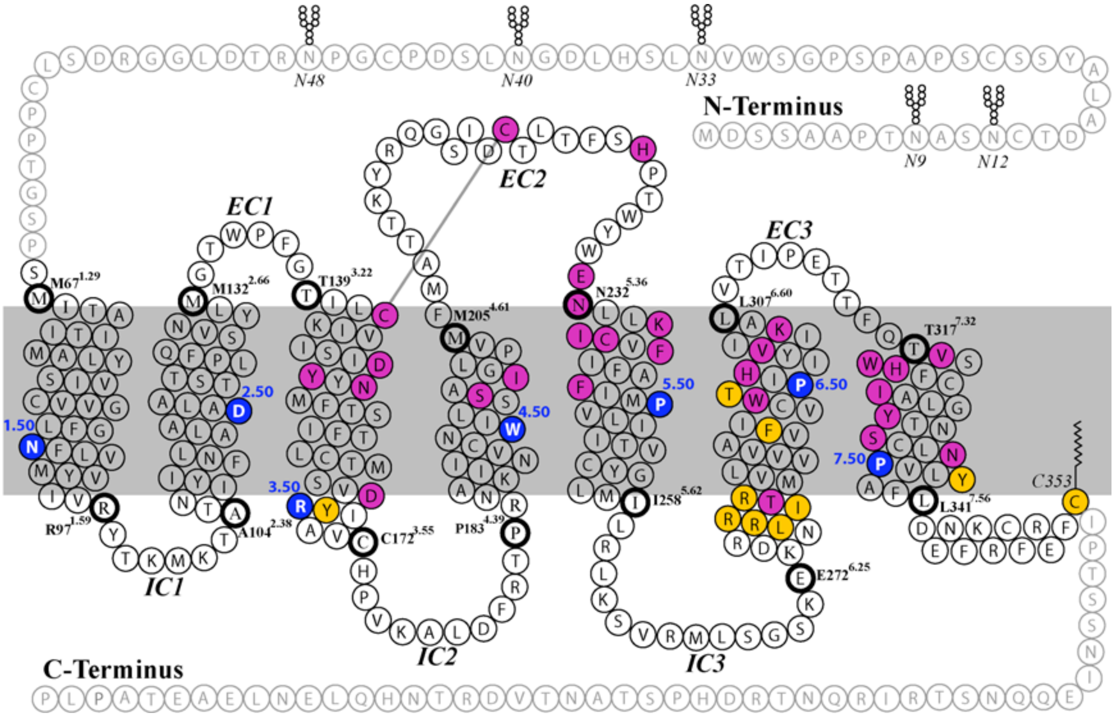


Figure 4.4. Snake-like Plot of the Entire Sequence of hMOP-R. The glycosylation sites are those proposed at UniProtKB [30]. The most conserved residues at each of the transmembrane helices are in blue circles with the respective Ballesteros and Weinstein indexing number. Positions involved in experimental studies mentioned in Table 4.1 are colored in magenta. The conserved motifs in the rhodopsin-like GPCR family are colored in orange. Disulfide bond between C142 and C219 is indicated. The N- and C- helical termini of each of the seven helices are depicted as bold circles. The membrane region is colored in gray and was obtained by aligning the 4T-hMOP-R model to the structure of the β_2 adrenergic receptor positioned in a membrane at the Orientation of Proteins in Membrane Database [80]. The absent (nonmodeled) sequence region in the model is display in gray circles.

4.3.5 Comparison with Previous Opioid Models

Three previously published models of the inactive conformation were compared with 4T-hMOP-R and 2T-hMOP-R. i) The human δ opioid receptor proposed by Kieffer and coworkers (Model_A) [81], ii) the model of human μ opioid receptor developed by Ferguson and coworkers (Model_B) [15], and iii) the rat μ opioid receptor model developed by Mosberg and coworkers (Model_C) [17]. There is a fourth model of the human μ opioid receptor but the coordinates are not available [19].

Model_A is based on the structure of bovine rhodopsin (PDB accession code: 1F88) and lacks the EC2 loop, which is important for ligand selectivity. The model contains the α-aneurysm in TM2 and therefore residue Thr122 is highly exposed to the lipid membrane. The volume for the binding cavity using CASTp with a radius probe of 1.4 Å was 1219.6 $Å^3$. The main differences between this model and 4T-hMOP-R and 2T-hMOP-R are the positions of the beginning of TM1 relative to the rest of the bundle, the presence of an α-aneurysm in TM2 and finally, the absence of the second extracellular loop EC2.

Model_B is based on electron cryomicroscopy data[82] and the C_α coordinate template developed by Baldwin *et al*. [83] This model comprises only the seven transmembrane helices. The TM2 helix does not contain the α-aneurysm. The main difference between this model and the 4T-hMOP-R and 2T-hMOP-R models is the presence of all the connecting loops in our models of hMOP-R, which are fundamental to address the receptor/ligand interactions.

Model_C is based on the structure of bovine rhodopsin (PDB accession code: 1GZM) and was refined using experimentally determined structural-distance constraints [17]. This model contains all the connecting loops, especially the important EC2 loop that forms a β-sheet structure and is buried in a similar way as the EC2 in bovine rhodopsin. In contrast, the 4T-hMOP-R and 2T-hMOP-R models have an EC2 loop that resembles the short helical segment present in the adrenergic receptors. This extracellular region adopts a more protruded position in comparison with the tightly packed loop of rhodopsin, leaving the entrace of the putative binding site directly exposed to water. Similar to Model_B, Model_C does not contain the α-aneurysm in TM2 and the first segment of the TM1 (residues Ser66$^{1.28}$ to Tyr77$^{1.39}$) is relatively closer to the rest of the helical bundle when compared with 4T-hMOP-R and 2T-hMOP-R. The volume for the binding cavity using CASTp with a radius probe of 1.4 Å was 1382.8 Å^3.

The models presented here (4T-hMOP-R and 2T-hMOP-R) provide information about the extracellular regions of the hMOP-R, in particular the structure of the EC2 loop. Only Model_C provides a structure of this important segment that is believed to be important for ligand affinity, selectivity and functional differences between hMOP-R and other receptors [1,7]. Model_C predicts an EC2 loop forming a short β-sheet structure that protects the ligand-binding pocket from solvent access as the EC2 does in rhodopsin. In the case of hMOP-R, and all the other available GPCR structures except rhodopsin, the endogenous ligands are diffusible and noncovalently bound. The structures of the EC2 loop display a more open binding site in comparison with rhodopsin, allowing access to the ligand-binding pocket.

The models presented here display significant improvements in the understanding of the structure of the hMOP-R when compared with the previously published models. They are able to aid in the interpretation of all the experimental data collected in Table 4.1. The models contain information about the extracellular loop (EC2) that connects helices TM4 and TM5, and the proposed structure for the EC2 loop is consistent with the experimental data [70-72]. In addition, the EC2 loop structure is consistent with the observation that other available GPCR structures having diffusible endogenous ligands present a more protruded structure when compared with that of rhodopsin allowing easier access to the ligand binding site [23-25].

4.3.6 Docking Calculations

Opioid ligands were computationally docked into the model structures. The crystal structure of β_2 adrenergic receptor [23], β_1 adrenergic receptor [24], and A_{2A} adenosine receptor [25] were solved with diffusible antagonists (partial inverse agonists in the case of β_2), while the structure of the bovine rhodopsin was solved with the covalently bound cofactor retinal in its 11-*cis* configuration [41], that is, all templates were solved in the inactive conformation. For this reason, an opioid antagonist (naloxone) was primarily utilized in the docking calculations to the μ opiod receptor models. Naloxone has high affinity for the μ opioid receptor and contains a positively charged nitrogen atom that is thought to strongly interact with the opioid receptor [6]. A classic opioid agonist, morphine, with a similar structure was also considered.

To test the reliability of the docking results, the four ligands in each of the template structures were docked into their corresponding receptors. In each case, the top docking results were in aggrement with the crystal structure even in the case of larger and more flexible ligands such as retinal and ZMA241385. The rmsd values between the crystallographic ligand positions and the docking results for the heavy atoms are: carazolol 1.49 Å, retinal 1.41 Å, cyanopindolol 0.83 Å and ZMA241385 2.69 Å.

Due to the lack of flexibility intrinsic in the docking calculations used here, modifications in the conformation of some side chains were needed to dock naloxone into 4T-hMOP-R and 2T-hMOP-R. Specifically, these involve changes in some dihedral angles of large residues such as Ile146, Tyr150, Trp228, Trp320, Ile324, and Tyr328 in the putative binding pocket. For instance, in the case of Trp228 the dihedral angle (χ) between N-C_α-C_β-C_γ was modified by 3.5° (from a value of -50.5° to 54.0°). All the structures were docked in a similar binding pocket as illustrated in Figure 4.5a. The binding site is located mainly between helices TM2, TM3, TM6 and TM7 with small involvement of TM5.

One of the docking results orients the positively charged ammonium nitrogen of naloxone in proximity to Asp149$^{3.32}$ in helix TM3 (Figure 4.5b and 4.5c). The interaction of opioid ligands with Asp149$^{3.32}$ has been found to be crucial for binding, and the mutation of this position leads to a substantial decrease in binding affinity [55,84]. Furthermore, the two adrenergic receptors have an aspartic acid located in an equivalent position (Asp113$^{3.32}$ for the β_2 and Asp121$^{3.32}$ for the β_1). In the crystal structure of β_2 and β_1 adrenergic receptors, both ligands (carazolol and cyanopindolol, respectively) contain a positively charged nitrogen atom interacting with this acidic residue. Figure 4.5d and 4.5e show a structural alignment of β_2 adrenergic receptor and carazolol (cyan) with the hMOP-R models and naloxone (gray and purple), where the positions of the positively charged nitrogens in carazolol and naloxone are similarly located. When the crystal structure of β_1 adrenergic receptor is used in the comparison, similar relative positioning of ligand's nitrogen atom is observed. The same models were used to dock the opioid agonist morphine. In both cases, morphine is positioned in the same binding cavity as naloxone, with orientations that bring the ammonium nitrogen of morphine in proximity to the acidic residue Asp149$^{3.32}$.

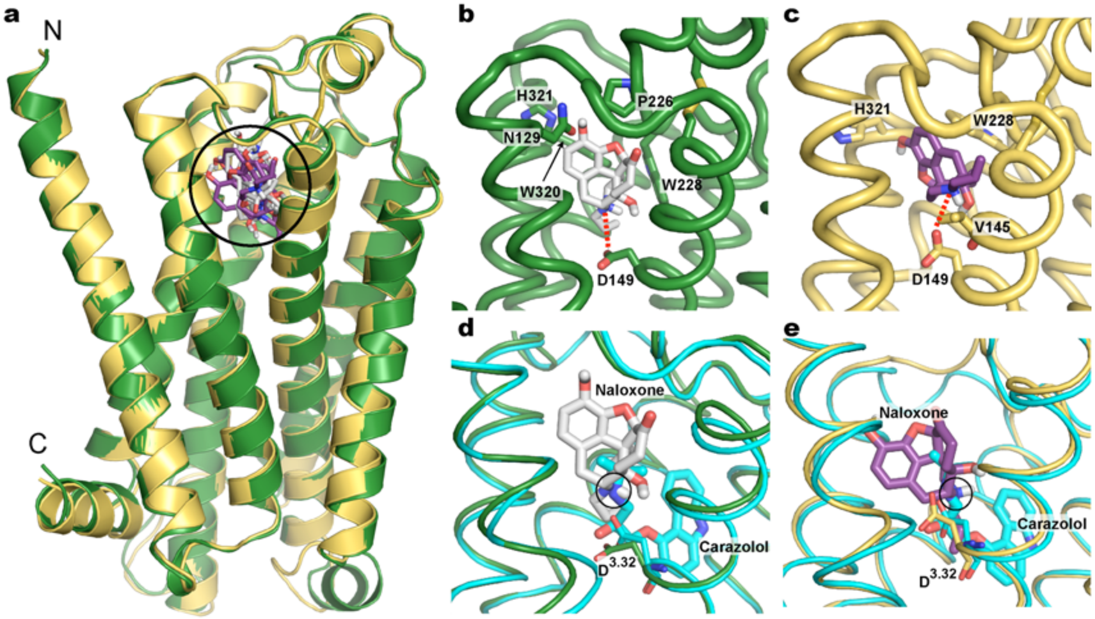


Figure 4.5. Docking Calculation Results for the Model Structures of hMOP-R. (**a**) Aligned structures of the 4T-hMOP-R (green) and 2T-hMOP-R (yellow) models with the top docked structures. The ligand molecules naloxone and morphine are located in the same binding site. The docked structures in the 4T-hMOP-R model are colored in grey while the docked structures in the 2T-hMOP-R model are colored in purple. (**b**) (**c**) Docked structure (naloxone) in a position that orients the positively charged nitrogen atom close to the oxygen atoms of the acidic residue Asp149$^{\mathbf{3.32}}$ (see dashed red line) for (**b**) 4T-hMOP-R and (**c**) 2T-hMOP-R. d, e, Superposition of crystal structure of β_2 adrenergic receptor with carazolol (cyan) with the (**d**) 4T-hMOP-R and (**e**) 2T-hMOP-R models. Carazolol's ammonium nitrogen atom (black circle) that interacts with residue Asp113$^{\mathbf{3.32}}$ that is analogous to Asp149$^{\mathbf{3.32}}$ in hMOP-R.

4.3.7 Residues Involved in Agonist/Antagonist Ligand Binding

Based on recent findings involving GPCR agonist-induced activation mechanisms [85-89], the hMOP-R models may be helpful in understanding the activation process of this receptor. Based on comparisons of recent structures of GPCRs bound to agonist and partial agonist [85-89], with their counterparts bound to antagonists and inverse agonists [23-25], residues have been identified to interact particularly with agonist and proposed to be important in agonist-induced activation.

In the case of the adrenergic receptors, β_1 and β_2, the important residues involved in agonist activation are located in TM5 [85,87]. Residues Ser211$^{5.42}$ and Ser215$^{5.46}$ form hydrogen bonds with both full and partial agonists, but just the full agonists interact with Ser215$^{5.46}$. The equivalent position of h-MOP-R contains an alanine residue Ala242$^{5.46}$.

In the case of the A_{2A} adenosine receptor, the residues involved in direct interaction with agonists and not with inverse agonists are located mainly in TM3 and TM7 [88,89]. Agonists formed non-polar interactions with Val84$^{3.32}$, Leu85$^{3.33}$ and Thr88$^{3.36}$ and polar interactions with Ser277$^{7.42}$ and His278$^{7.43}$. Both kinds of interactions were concluded to be important in the activation of the A_{2A} adenosine receptor [88,89].

The general receptor structural changes upon agonist-induced activation were concluded to be: i) The inward motion of TM3, TM5 and TM7 (contraction of the binding pocket), and ii) the formation of a bulge in TM5 (centered around Ser215$^{5.46}$ and Ser207$^{5.46}$ in

β_1 and β_2 adrenergic receptors, and Cys185$^{5.46}$ in the adenosine receptor) [89]. Finally, based on structural alignments, it was concluded that the positions involved in contact with agonist in the complexes β_1 adrenergic receptor/isoprenaline and A_{2A} adenosine receptor/adenosine are: Asp121$^{3.32}$, Val84$^{3.32}$; Val125$^{3.36}$, Thr88$^{3.36}$; Ser211$^{5.42}$, Asn181$^{5.42}$; Phe306$^{6.51}$, Leu249$^{6.51}$; Asn310$^{6.55}$, Asn253$^{6.55}$; Phe307$^{6.52}$, His250$^{6.52}$; Asn329$^{7.39}$, Ile274$^{7.39}$ and Tyr333$^{7.43}$, His278$^{7.43}$, respectively (see Figure 4.6a) [89]. Using the representative models of the human μ opioid receptor, 4T-hMOP-R and 2T-hMOP-R, the equivalent positions are Asp149$^{3.32}$, Met153$^{3.36}$, Val238$^{5.42}$, Ile298$^{6.51}$, Val302$^{6.55}$, His299$^{6.52}$, Ile324$^{7.39}$ and Tyr328$^{7.43}$. To our knowledge, there have been studies on hMOP-R involving Asp149$^{3.32}$, Val302$^{6.55}$, His299$^{6.52}$, Ile324$^{7.39}$ and Tyr328$^{7.43}$. The effects of mutations involving residues Met153$^{3.36}$, Val238$^{5.42}$, and Ile298$^{6.51}$ have yet to be explored, however, particularly with regard to agonist-induced activation in μ opioid receptor, and would be interesting candidates for future studies.

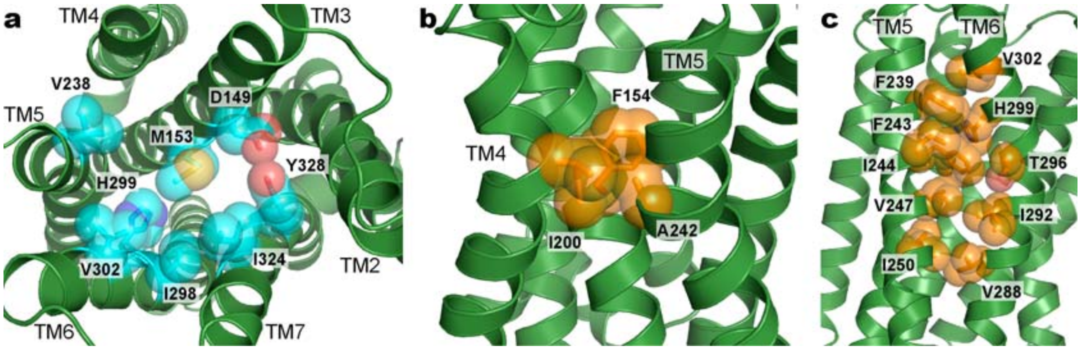


Figure 4.6. Residues Involve in Agonist-Induced Activation of hMOP-R. (**a**) Equivalent positions in the hMOP-R that make contacts with agonists in the complexes A_{2A} adenosine receptor/adenosine and β_1 adrenergic receptor/isoprenaline (cyan space-filling models).[89] To our knowledge, positions Met153$^{\mathbf{3.36}}$, Val238$^{\mathbf{5.42}}$, and Ile298$^{\mathbf{6.51}}$ have not been investigated to address their relevance in agonist binding. (**b**) Ala242$^{\mathbf{5.46}}$ is centered in the bulge formation upon agonist binding. (**c**) The interface of TM5 and TM6 is formed mostly by hydrophobic interactions. In order to reinforce the interaction of TM5 and TM6, mutation of position Thr296$^{6.44}$ for a more hydrophobic residue is anticipated to have similar effects as the hydrogen bond formation between Ser$^{\mathbf{5.43}}$ and Asn$^{\mathbf{6.55}}$ in the agonist bound adrenergic receptors.

From these previous structural studies of the adenosine/agonist and adrenergic/agonist complexes and the structural information from the 4T-hMOP-R and the 2T-hMOP-R models, two main predictions for the hMOP-R are suggested:

1) Given the bulge formation in TM5 upon agonist binding and the thermostabilizing mutations in the A_{2A} adenosine receptor, there may be an important interaction between positions Cys185$^{5.46}$ and Gln89$^{3.37}$ [88,89]. These positions have been shown to be relevant to agonist binding.[85] The mutation Q89A stabilizes the active state of the A_{2A} adenosine receptor,[85] so disruption of the interaction at this position in hMOP-R is suggested to be also relevant (the equivalent positions in h-MOP-R are Ala242$^{5.46}$ and Phe154$^{3.37}$, see Figure 4.6b). Based on the models of h-MOP-R, mutation of Ala242$^{5.46}$ could have significant implications in the activation of the receptor, Moreover, position Ile200$^{4.56}$ also seems to play an important role in these interactions (see Figure 4.6b). Mutagenesis of this position for a smaller residue I200V, decreases the agonist affinity (K_d) of morphine and the synthetic opioid peptide DAMGO by a factor of 4 to 5 [57]. Using the models proposed here, the explanation for this result is that a smaller residue reduces the propensity toward inward movement of residue Ala242$^{5.46}$ (center of the bulge formation upon agonist binding), thus we anticipate that a bulkier residue would facilitate the inward movement of Ala242$^{5.46}$, shifting the equilibrium to the active state and therefore increasing agonist binding affinity.

2) Upon binding of agonists in the case of the adrenergic receptors, there is a formation of a hydrogen bond (Ser212$^{5.43}$ and Asn310$^{6.55}$ in β_1, Ser204$^{5.43}$ and Asn293$^{6.55}$ in β_2) between helices TM5 and TM6. In the case of hMOP-R, the residue identities at these positions (Phe239$^{5.43}$ and Val302$^{6.55}$) prevent similar hydrogen binding interactions. If the reinforcement of the interactions of helices TM5 and TM6 occurs upon agonist binding, using the models presented here, mutation of Thr296$^{6.44}$ for a more hydrophobic residue could strengthen the interactions of TM5 and TM6 in a similar fashion to that observed in the β_1 (β_2) adrenergic

receptors. As a result, a shift in the equilibrium to active conformations and an increase in agonist binding would be expected (see Figure 4.6c).

4.4 Conclusions

Models of the human μ opioid receptor have been created that leverage new structural data concerning GPCRs. We compare varying the number of templates used in the comparative modeling. Interestingly, the ensemble analysis showed that there is small difference when four templates and two templates are used to create the models. One of the main conclusions is that both ensembles produced similar results on the condition that the sequence alignment used is the same. More templates usually produce better sequence coverage and therefore the hypothesis that more templates are able to improve the comparative models is still reasonable. We have seen from the different GPCR structures available that they have some structural diversity and therefore surmount some of the structural bias inherent in using just one structure. Nonetheless, the structures used in this study are still limited to the four mentioned herein.

Based on the energies after minimization, two models were selected as representative of each ensemble. Detailed analysis of 4T-hMOP-R and 2T-hMOP-R shows that the two models are similar and that both may be used to interpret the collected experimental data. One important difference is the second intracellular loop (IC2) that connects TM3 and TM4. For this segment, 4T-hMOP-R forms a helical motif that is inherited from the β_1 adrenergic receptor and the A_{2A} adenosine receptor and therefore is absent in the 2T-hMOP-R model. This region of the receptor is relevant because it is close to the highly conserved *DRY* motif.

The docking calculations were in aggrement with some experimental data and they show a possible binding orientation of the ligand. Also, both models produce similar results for the docking of an antagonist and an agonist. Naloxone, an opioid antagonist, was docked into 4T-hMOP-R and 2T-hMOP-R. One particular docking result, oriented the positively charged nitrogen of naloxone, close to the acidic residue Asp149$^{3.32}$ located in helix TM3. This residue has been shown to be crucial for the binding of different ligands. The crystal structures of β_2 and β_1 adrenergic receptors show that the analogous residue Asp$^{3.32}$ is interacting with a positively charged nitrogen atom from carazolol and cyanopindolol, respectively. When the structures of 4T-hMOP-R/naloxone and 2T-hMOP-R/naloxone are superimposed with the structure of β_2 adrenergic/carazolol and β_1 adrenergic receptor/cyanopindolol, the positions of the positively charged nitrogens are almost identical. The opioid agonist morphine was docked in the models and binds in the same binding cavity as naloxone, with the positively charged nitrogen of morphine close the residue Asp149$^{3.32}$. Based on information from GPCRs agonist bound structures and the suggested agonist-induced mechanisms, predictions about residues involved in the agonist-induced activation process in the hMOP-R are anticipated. Specifically, the following is proposed. (1) A substitution that facilitates the inward movement of TM5 and the bulge formation centered at Ala242$^{5.46}$. The mutation involves interaction with a bulkier residue at position Ile200$^{4.56}$. (2) A single mutation to strengthen the interactions between TM5 and TM6 in hMOP-R, based on the suggested reinforced interactions upon agonist binding between TM5 and TM6. The substitution involves a hydrophobic residue at position Thr296$^{6.44}$.

4.5 References

1. Kieffer BL, Evans CJ: **Opioid receptors: From binding sites to visible molecules in vivo**. *Neuropharmacology* 2009, **56**:205-212.
2. Corbett AD, Henderson G, McKnight AT, Paterson SJ: **75 years of opioid research: the exciting but vain quest for the Holy Grail**. *British Journal of Pharmacology* 2006, **147**:S153-S162.
3. Waldhoer M, Bartlett SE, Whistler JL: **Opioid receptors**. *Annual Review of Biochemistry* 2004, **73**:953-990.

4. Martin M, Matifas A, Maldonado R, Kieffer BL: **Acute antinociceptive responses in single and combinatorial opioid receptor knockout mice: distinct mu, delta and kappa tones**. *Eur J Neurosci* 2003, **17**:701-708.

5. Bodnar RJ: **Endogenous opiates and behavior: 2006**. *Peptides* 2007, **28**:2435-2513.

6. Goodman AJ, Le Bourdonnec B, Dolle RE: **Mu opioid receptor antagonists: recent developments**. *ChemMedChem* 2007, **2**:1552-1570.

7. Rosenbaum DM, Rasmussen SG, Kobilka BK: **The structure and function of G-protein-coupled receptors**. *Nature* 2009, **459**:356-363.

8. Pierce KL, Premont RT, Lefkowitz RJ: **Seven-transmembrane receptors**. *Nat Rev Mol Cell Biol* 2002, **3**:639-650.

9. Hanson MA, Stevens RC: **Discovery of new GPCR biology: one receptor structure at a time**. *Structure* 2009, **17**:8-14.

10. Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB: **The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints**. *Mol Pharmacol* 2003, **63**:1256-1272.

11. Lagerstrom MC, Schioth HB: **Structural diversity of G protein-coupled receptors and significance for drug discovery**. *Nature Reviews Drug Discovery* 2008, **7**:339-357.

12. Moles A, Kieffer BL, D'Amato FR: **Deficit in attachment behavior in mice lacking the mu-opioid receptor gene**. *Science* 2004, **304**:1983-1986.

13. Contet C, Kieffer BL, Befort K: **Mu opioid receptor: a gateway to drug addiction**. *Curr Opin Neurobiol* 2004, **14**:370-378.

14. Filizola M, Carteni-Farina M, Perez JJ: **Molecular modeling study of the differential ligand-receptor interaction at the mu, delta and kappa opioid receptors**. *Journal of Computer-Aided Molecular Design* 1999, **13**:397-407.

15. Subramanian G, Paterlini MG, Portoghese PS, Ferguson DM: **Molecular docking reveals a novel binding site model for fentanyl at the mu-opioid receptor**. *Journal of Medicinal Chemistry* 2000, **43**:381-391.

16. Zhorov BS, Ananthanarayanan VS: **Homology models of mu-opioid receptor with organic and inorganic cations at conserved aspartates in the second and third transmembrane domains**. *Archives of Biochemistry and Biophysics* 2000, **375**:31-49.

17. Fowler CB, Pogozheva ID, LeVine H, Mosberg HI: **Refinement of a homology model of the mu-opioid receptor using distance constraints from intrinsic and engineered zinc-binding sites**. *Biochemistry* 2004, **43**:8700-8710.

18. Zhang Y, Sham YY, Rajamani R, Gao J, Portoghese PS: **Homology modeling and molecular dynamics simulations of the mu opioid receptor in a membrane-aqueous system**. *Chembiochem* 2005, **6**:853-859.

19. Bera I, Laskar A, Ghoshal N: **Exploring the structure of opioid receptors with homology modeling based on single and multiple templates and subsequent docking: A comparative study**. *J Mol Model* 2010.

20. Baker D, Sali A: **Protein structure prediction and structural genomics**. *Science* 2001, **294**:93-96.

21. Jorgensen WL: **The many roles of computation in drug discovery**. *Science* 2004, **303**:1813-1818.

22. Klabunde T, Hessler G: **Drug design strategies for targeting G-protein-coupled receptors**. *Chembiochem* 2002, **3**:929-944.

23. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, et al.: **High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor**. *Science* 2007, **318**:1258-1265.

24. Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AG, Tate CG, Schertler GF: **Structure of a beta1-adrenergic G-protein-coupled receptor**. *Nature* 2008, **454**:486-491.

25. Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, Ijzerman AP, Stevens RC: **The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist**. *Science* 2008, **322**:1211-1217.

26. Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VR, Sanishvili R, Fischetti RF, et al.: **Crystal structure of the human beta2 adrenergic G-protein-coupled receptor**. *Nature* 2007, **450**:383-387.

27. Kane BE, Svensson B, Ferguson DM: **Molecular recognition of opioid receptor ligands**. *Aaps Journal* 2006, **8**:E126-E137.

28. Elofsson A, Larsson P, Wallner B, Lindahl E: **Using multiple templates to improve quality of homology models in automated homology modeling**. *Protein Science* 2008, **17**:990-1002.

29. Niv MY, Yarnitzky T, Levit A: **Homology modeling of G-protein-coupled receptors with X-ray structures on the rise**. *Current Opinion in Drug Discovery & Development* 2010, **13**:317-325.

30. Jain E, Bairoch A, Duvaud S, Phan I, Redaschi N, Suzek BE, Martin MJ, McGarvey P, Gasteiger E: **Infrastructure for the life sciences: design and implementation of the UniProt website**. *BMC Bioinformatics* 2009, **10**:136.

31. Michino M, Abola E, Brooks CL, Dixon JS, Moult J, Stevens RC, Participants GD: **Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008**. *Nature Reviews Drug Discovery* 2009, **8**:455-463.

32. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A: **Comparative protein structure modeling of genes and genomes**. *Annu Rev Biophys Biomol Struct* 2000, **29**:291-325.

33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**:403-410.

34. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A: **ExPASy: The proteomics server for in-depth protein knowledge and analysis**. *Nucleic Acids Res* 2003, **31**:3784-3788.

35. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al.: **Clustal W and Clustal X version 2.0**. *Bioinformatics* 2007, **23**:2947-2948.

36. Kelley LA, Sternberg MJ: **Protein structure prediction on the Web: a case study using the Phyre server**. *Nat Protoc* 2009, **4**:363-371.

37. Henikoff S, Henikoff JG: **Amino acid substitution matrices from protein blocks**. *Proc Natl Acad Sci U S A* 1992, **89**:10915-10919.

38. Schwartz RM, Dayhoff MO: **Improved scoring matrix for identifying evolutionary relatedness among proteins**. *Biophysical Journal* 1978, **21**:A198-A198.

39. Surratt CK, Adams WR: **G protein-coupled receptor structural motifs: relevance to the opioid receptors**. *Curr Top Med Chem* 2005, **5**:315-324.

40. Frishman D, Argos P: **Knowledge-based protein secondary structure assignment**. *Proteins* 1995, **23**:566-579.

41. Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V: **The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure**. *J Mol Biol* 2004, **342**:571-583.

42. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints**. *J Mol Biol* 1993, **234**:779-815.

43. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A: **Comparative protein structure modeling using Modeller**. *Curr Protoc Bioinformatics* 2006, **Chapter 5**:Unit 5 6.

44. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K: **Scalable molecular dynamics with NAMD**. *J Comput Chem* 2005, **26**:1781-1802.

45. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, et al.: **All-atom empirical potential for molecular modeling and dynamics studies of proteins**. *Journal of Physical Chemistry B* 1998, **102**:3586-3616.

46. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC: **MolProbity: all-atom structure validation for macromolecular crystallography**. *Acta Crystallogr D Biol Crystallogr* 2010, **66**:12-21.

47. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J: **CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues**. *Nucleic Acids Res* 2006, **34**:W116-118.

48. Hazai E, Kovacs S, Demko L, Bikadi Z: **[DockingServer: molecular docking calculations online]**. *Acta Pharm Hung* 2009, **79**:17-21.

49. Huey R, Morris GM, Olson AJ, Goodsell DS: **A semiempirical free energy force field with charge-based desolvation**. *J Comput Chem* 2007, **28**:1145-1152.

50. Bolton E, Wang Y, Thiessen PA, Bryant SH: **PubChem: Integrated platform of small molecules and biological activities**. *Annual Reports in Computational Chemistry* 2008, **4**.

51. Stewart JJP: **MOPAC2009 Stewart Computational Chemistry, Colorado Springs, CO, USA**. *OpenMOPAC.net* 2008.

52. Solis FJ, Wets RJB: **Minimization by random search techniques**. *Mathematics of Operations Research* 1981, **6**:19-30.

53. Hanson MA, Cherezov V, Griffith MT, Roth CB, Jaakola VP, Chien EY, Velasquez J, Kuhn P, Stevens RC: **A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor**. *Structure* 2008, **16**:897-905.

54. Heerding J, Raynor K, Kong H, Yu L, Reisine T: **Mutagenesis Reveals That Agonists and Peptide Antagonists Bind in Fundamentally Distinct Manners to the Rat Mu-Receptor Than Do Nonpeptide Antagonists**. *Regulatory Peptides* 1994, **54**:119-120.

55. Li JG, Chen CG, Yin JL, Rice K, Zhang Y, Matecka D, de Riel JK, DesJarlais RL, Liu-Chen LY: **Asp147 in the third transmembrane helix of the rat mu opioid receptor forms ion-pairing with morphine and naltrexone**. *Life Sciences* 1999, **65**:175-185.

56. Xu H, Lu YF, Partilla JS, Zheng QX, Wang JB, Brine GA, Carroll FI, Rice KC, Chen KX, Chi ZQ, et al.: **Opioid peptide receptor studies, 11: Involvement of Tyr148, Trp318 and His319 of the rat mu-opioid receptor in binding of mu-selective ligands**. *Synapse* 1999, **32**:23-28.

57. Mansour A, Taylor LP, Fine JL, Thompson RC, Hoversten MT, Mosberg HI, Watson SJ, Akil H: **Key residues defining the mu-opioid receptor binding pocket: A site-directed mutagenesis study**. *Journal of Neurochemistry* 1997, **68**:344-353.

58. Li J, Huang P, Chen CG, de Riel JK, Weinstein H, Liu-Chen LY: **Constitutive activation of the mu opioid receptor by mutation of D3.49(164), but not D3.32(147): D3.49(164) is critical for stabilization of the inactive form of the receptor and for its expression**. *Biochemistry* 2001, **40**:12039-12050.

59. Xu W, Sanz A, Pardo L, Liu-Chen LY: **Activation of the mu opioid receptor involves conformational rearrangements of multiple transmembrane domains**. *Biochemistry* 2008, **47**:10576-10586.

60. Claude PA, Wotta DR, Zhang XH, Prather PL, McGinn TM, Erickson LJ, Loh HH, Law PY: **Mutation of a conserved serine in TM4 of opioid receptors confers full agonistic properties to classical antagonists**. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**:5715-5719.

61. Pil J, Tytgat J: **The role of the hydrophilic Asn230 residue of the mu-opioid receptor in the potency of various opioid agonists**. *British Journal of Pharmacology* 2001, **134**:496-506.

62. Chen C, Yin J, Riel JK, DesJarlais RL, Raveglia LF, Zhu J, Liu-Chen LY: **Determination of the amino acid residue involved in [3H]beta-funaltrexamine covalent binding in the cloned rat mu-opioid receptor**. *J Biol Chem* 1996, **271**:21422-21429.

63. Huang P, Li J, Chen CG, Visiers I, Weinstein H, Liu-Chen LY: **Functional role of a conserved motif in TM6 of the rat mu opioid receptor: Constitutively active and inactive receptors result from substitutions of Thr6.34(279) with Lys and Asp**. *Biochemistry* 2001, **40**:13501-13509.

64. Surratt CK, Johnson PS, Moriwaki A, Seidleck BK, Blaschak CJ, Wang JB, Uhl GR: **-Mu Opiate Receptor - Charged Transmembrane Domain Amino-Acids Are Critical for Agonist Recognition and Intrinsic Activity**. *Journal of Biological Chemistry* 1994, **269**:20548-20553.

65. Jones RM, Hjorth SA, Schwartz TW, Portoghese PS: **Mutational evidence for a common kappa antagonist binding pocket in the wild-type kappa and mutant mu[K303E] opioid receptors**. *Journal of Medicinal Chemistry* 1998, **41**:4911-4914.

66. Metzger TG, Paterlini MG, Ferguson DM, Portoghese PS: **Investigation of the selectivity of oxymorphone- and naltrexone-derived ligands via site-directed mutagenesis of opioid receptors: Exploring the 'address' recognition locus**. *Journal of Medicinal Chemistry* 2001, **44**:857-862.

67. Ulens C, Van Boven M, Daenens P, Tytgat J: **Interaction of p-fluorofentanyl on cloned human opioid receptors and exploration of the role of Trp-318 and His-319 in mu-opioid receptor selectivity**. *Journal of Pharmacology and Experimental Therapeutics* 2000, **294**:1024-1033.

68. Owens CE, Akil H: **Determinants of ligand selectivity at the kappa-receptor based on the structure of the orphanin FQ receptor**. *Journal of Pharmacology and Experimental Therapeutics* 2002, **300**:992-999.

69. Pil J, Tytgat J: **Serine 329 of the mu-opioid receptor interacts differently with agonists**. *Journal of Pharmacology and Experimental Therapeutics* 2003, **304**:924-930.

70. Zhang PS, Johnson PS, Zollner C, Wang WF, Wang ZJ, Montes AE, Seidleck BK, Blaschak CJ, Surratt CK: **Mutation of human mu opioid receptor extracellular "disulfide cysteine" residues alters ligand binding but does not prevent receptor targeting to the cell plasma membrane**. *Molecular Brain Research* 1999, **72**:195-204.

71. Shahrestanifar M, Wang WW, Howells RD: **Studies on inhibition of mu and delta opioid receptor binding by dithiothreitol and N-ethylmaleimide - His223 is critical for mu opioid receptor binding and inactivation by N-ethylmaleimide**. *Journal of Biological Chemistry* 1996, **271**:5505-5512.

72. Fowler CB, Pogozheva ID, Lomize AL, LeVine H, Mosberg HI: **Complex of an active mu-opioid receptor with a cyclic peptide agonist modeled from experimental constraints.** *Biochemistry* 2004, **43**:15796-15810.

73. Xu W, Ozdener F, Li JG, Chen CG, de Riel JK, Weinstein H, Liu-Chen LY: **Functional role of the spatial proximity of Asp114(2.50) in TMH2 and Asn332(7.49) in TMH7 of the mu opioid receptor**. *Febs Letters* 1999, **447**:318-324.

74. Zhang Y, McCurdy CR, Metzger TG, Portoghese PS: **Specific cross-linking of lys233 and cys235 in the mu opioid receptor by a reporter affinity label**. *Biochemistry* 2005, **44**:2271-2275.

75. Befort K, Tabbara L, Kling D, Maigret B, Kieffer BL: **Role of aromatic transmembrane residues of the delta-opioid receptor in ligand recognition**. *Journal of Biological Chemistry* 1996, **271**:10161-10168.

76. Seki T, Minami M, Nakagawa T, Ienaga Y, Morisada A, Satoh M: **DAMGO recognizes four residues in the third extracellular loop to discriminate between mu- and kappa-opioid receptors**. *European Journal of Pharmacology* 1998, **350**:301-310.

77. Ballesteros JA, Weinstein H: **Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors**. *Methods Neurosci.* 1995, **25**:366-428.

78. Vogel R, Mahalingam M, Luedke S, Huber T, Siebert F, Sakmar TP: **Functional role of the "Ionic Lock" - An interhelical hydrogen-bond network in family a heptahelical receptors**. *Journal of Molecular Biology* 2008, **380**:648-655.

79. Bond RA, Ijzerman AP: **Recent developments in constitutive receptor activity and inverse agonism, and their potential for GPCR drug discovery**. *Trends Pharmacol Sci* 2006, **27**:92-96.

80. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI: **OPM: Orientations of proteins in membranes database**. *Bioinformatics* 2006, **22**:623-625.

81. Decaillot FM, Befort K, Filliol D, Yue SY, Walker P, Kieffer BL: **Opioid receptor random mutagenesis reveals a mechanism for G protein-coupled receptor activation**. *Nature Structural Biology* 2003, **10**:629-636.

82. Unger VM, Hargrave PA, Baldwin JM, Schertler GF: **Arrangement of rhodopsin transmembrane alpha-helices**. *Nature* 1997, **389**:203-206.

83. Baldwin JM, Schertler GF, Unger VM: **An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors**. *J Mol Biol* 1997, **272**:144-164.

84. Befort K, Tabbara L, Bausch S, Chavkin C, Evans C, Kieffer B: **The conserved aspartate residue in the third putative transmembrane domain of the delta-opioid receptor is not the anionic counterpart for cationic opiate binding but is a constituent of the receptor binding site**. *Molecular Pharmacology* 1996, **49**:216-223.

85. Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, Arlow DH, Rasmussen SGF, Choi HJ, DeVree BT, Sunahara RK, et al.: **Structure and function of an irreversible agonist-beta(2) adrenoceptor complex**. *Nature* 2011, **469**:236-240.

86. Rasmussen SGF, Choi HJ, Fung JJ, Pardon E, Casarosa P, Chae PS, DeVree BT, Rosenbaum DM, Thian FS, Kobilka TS, et al.: **Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor**. *Nature* 2011, **469**:175-180.

87. Warne T, Moukhametzianov R, Baker JG, Nehme R, Edwards PC, Leslie AGW, Schertler GFX, Tate CG: **The structural basis for agonist and partial agonist action on a beta(1)-adrenergic receptor**. *Nature* 2011, **469**:241-244.

88. Xu FX, F., Wu HX, Katritch V, Han GW, Jacobson KA, Gao ZG, Cherezov V, Stevens RC: **Structure of an Agonist-Bound Human A(2A) Adenosine Receptor**. *Science* 2011, **332**:322-327.

89. Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, Tate CG: **Agonist-bound adenosine A(2A) receptor structures reveal common features of GPCR activation**. *Nature* 2011, **474**:521-U154.

5 Computational Protein Design and Molecular Dynamics Simulations of Stromal Cell–Derived Factor–1α and Simplified Peptide Analogs[1]

5.1 Computational Protein Design to Re-Engineer Stromal Cell–Derived Factor–1α Generates an Effective and Translatable Angiogenic Polypeptide Analog[2]

5.1.1 Introduction

More people die annually of cardiovascular disease than of any other cause. The World Health Organization estimates that more than 17 million people per year die of cardiovascular disease. Of these deaths, an estimated 7.2 million were due to ischemic heart disease. An estimated 80 000 000 American adults (approximately 1 in 3) have at least 1 type of cardiovascular disease, with more than 22 000 000 having either coronary heart disease or heart failure [1].

---

[1] This project was a collaboration with the group of Professor Y. Joseph Woo at Division of Cardiovascular Surgery, Department of Surgery, University of Pennsylvania School of Medicine, Philadelphia, PA 19104.

[2] Adapted in part from William Hiesinger, Jose Manuel Perez-Aguilar, Pavan Atluri, Nicole A. Marotta, John R. Frederick, J. Raymond Fitzpatrick III, Ryan C. McCormick, Jeffrey R. Muenzer, Elaine C. Yang, Rebecca D. Levit, Li-Jun Yuan, John W. MacArthur, Jeffery G. Saven, Y. Joseph Woo *Circulation* **2011,** 124:S18-26.

Treatments for ischemic heart disease often fail because they do not address the underlying ventricular cellular pathophysiology and, most importantly, do not restore microvascular perfusion, which has been shown to be a critical, independent predictor of ventricular remodeling, as well as reinfarction, heart failure, and death [2,3]. On the other hand, postinfarction patients who develop robust angiographic collateralization manifest improved regional ventricular function, which suggests an important role for the native revascularization process [4]. Current therapeutic options for ischemic heart disease are only capable of intervening in relatively large arteries, which leaves pervasive microvascular dysfunction unaddressed. This is important because microvascular integrity, specifically microvascular blood velocity and flow, predicts functional recovery of ischemic myocardium [5-9].

In addition, current interventions are instituted relatively late in the overall time course of the disease process, with macrorevascularization possible in only 63% to 80% of patients with ischemic heart disease [10]. Even with reestablishment of epicardial coronary artery flow through thrombolysis, percutaneous intervention, or bypass grafting, a paucity of patent, functional microvasculature remains. Angiogenic cytokine therapy is a microrevascularization strategy that can serve as a primary therapy at any point in the disease process and can be used synergistically with traditional coronary revascularization methods [11]. A wide variety of cytokines and microrevascularization techniques have been used with varying degrees of success [12-15].

Stromal cell–derived factor–1α (SDF) is a powerful chemoattractant and is considered to be one of the key regulators of hematopoietic stem cell trafficking between the peripheral circulation and bone marrow. It has been shown to effect proliferation and mobilization of endothelial progenitor cells (EPCs) to induce vasculogenesis and is significantly upregulated in response to both myocardial ischemia and infarction [16,17]. SDF, a 67–amino acid protein, is also remarkably conserved among species; a single amino acid substitution is all that differentiates the human and murine sequences [18]. Experimentally, delivery of SDF to ischemic myocardium has been shown to increase circulating EPCs, significantly enhance myocardial EPC density, increase vasculogenesis, and augment myocardial function by enhancing perfusion, reversing cellular ischemia, increasing cardiomyocyte viability, and preserving ventricular geometry [19,20].

However, exogenous SDF is quickly degraded by multiple proteases [18,21], has a large and complex tertiary structure that involves multiple disulfide linkages [22], and is very expensive because it is a recombinant protein. Smaller analogs of SDF may provide translational advantages, including enhanced stability and function, ease of synthesis, lower cost, and potential modulated delivery via engineered biomaterials [23,24]. In the present study, we set out to design and engineer a minimized, highly efficient polypeptide analog of the SDF molecule using computational molecular modeling and design. The primary design goal was to remove the large, central β-sheet region and link the native N–terminus (responsible for receptor activation and binding) and the C–terminus (responsible for extracellular stabilization) while maintaining functionality and the approximately perpendicular orientation between the 2 termini [22]. Two disulfide bonds located in the large central β-sheet region help to maintain the conformation of native SDF. To recover a similar conformation after excision of this portion of the protein, proline residue linkers were considered to connect the N– and C–termini. The cyclic proline residue limits the conformational space accessible to the polypeptide. Loop modeling calculations revealed the variability of conformations explored by linkers comprising different numbers of proline residues. A designed diproline linker yielded energetic and conformational advantages that resulted in a small, low-molecular-weight engineered SDF polypeptide analog (ESA) that was shown to have activity comparable to or better than recombinant human SDF both in vitro and in a murine model of ischemic heart failure.

5.1.2 Methods

5.1.2.1 Computational Protein Design, Modeling, and Synthesis

Loop modeling calculations were performed with the program MODELLER, which implements a loop modeling algorithm that consists of an optimization of a defined segment of protein structure in a fixed environment, guided by a pseudo–energy-scoring function [25]. Fifty different structures for a 1-, 2-, and 3-proline residue linker were created while constraining the coordinates of the remainder of the protein. The conformation with the lowest effective energy (best score) was analyzed with MolProbity, which provides detailed all-atom contact analysis of steric interactions, dihedral angles, and possible hydrogen bonds and van der Waals contacts [26]. After the amino acid sequence of the newly designed ESA peptide had been computationally designed and analyzed, it was synthesized by solid-phase peptide synthesis, which involves the incorporation of N α-amino acids into a peptide of any desired sequence with 1 end of the sequence remaining attached to a solid support matrix. After the desired sequence of amino acids has been obtained, the peptide is removed from the polymeric support.

5.1.2.2 Cell Isolation

Bone marrow mononuclear cells were isolated from the long bones of adult male green fluorescent protein (GFP)-expressing Wistar rats by density-gradient centrifugation with Histopaque 1083 (Sigma-Aldrich) and cultured in endothelial basal medium-2 supplemented with EGM-2 SingleQuot (Lonza) containing human epidermal growth factor, FBS, vascular endothelial growth factor, basic human fibroblast growth factor, recombinant human long R3 insulin-like growth factor-1, ascorbic acid, heparin, gentamicin, and amphotericin-B. The combination of endothelium-specific media and the removal of nonadherent bone marrow mononuclear cells was intended to select for the EPC phenotype.

5.1.2.3 Functional Characterization of Isolated EPCs by Boyden Chamber Assay

Boyden chambers (Neuro Probe) were used to assess EPC migration. Briefly, 8-μm filters were loaded into control and experimental chambers. Fourteen-day EPCs cultured in endothelium-specific media on vitronectin-coated plates were trypsinized, counted, and brought to a concentration of 90 cells/μL in Dulbecco phosphate buffered saline (DPBS). The bottom chamber of the control and experimental chambers were loaded with DPBS, 100 ng/mL recombinant SDF (R&D Systems) in DPBS, or 100 ng/mL ESA in DPBS. A 560-μL cell suspension was added to the top chamber of each. All 3 chambers were incubated at 37°C, 5% CO2 for 4 hours. The cells remaining in the top chamber were wiped clean with a cotton swab, and the filter was removed. Slides were visualized on a DF5000B Leica Fluorescent scope and analyzed via LASAF version 2.0.2 (Leica) software. Boyden chamber analysis was performed in triplicate. Additional Boyden chamber assays were run to generate a dose-response curve to compare molar equivalents of ESA and SDF. A series of progressively increasing ESA and SDF gradients (0.625, 1.25, 2.5, 5, 12.5, 20, 25, and 50 nmol/L) were used, and assays at each dose were run 10 times.

5.1.2.4 CXCR4 Receptor Activation Assay

Fourteen-day EPCs cultured in endothelium-specific media on vitronectin-coated plates were trypsinized and plated onto the bottom of a 96-well cell culture plate at a density of 10 000 cells per well and allowed to adhere for 24 hours. Cells were incubated with culture media that contained recombinant SDF (0.5 μg/mL) or ESA (0.5 μg/mL) or media only (control) for 10 minutes and subsequently were fixed in 4% formaldehyde. Phosphorylated and total AKT levels were quantified with the FACE AKT ELISA kit (Active Motif). These experiments were also repeated but with the addition of the chemokine receptor 4 (CXCR4) receptor antagonist AMD3100 to each experimental group. Results are reported in optical density (OD) units.

5.1.2.5 CXCR4 Receptor Expression

EPCs were seeded overnight at a density of 1×104 cells per coverslip in 4-well slide chambers. Cells were either treated with 100 ng/mL SDF or 100 ng/mL ESA at 37°C for 6 hours. After chemokine stimulation, the cells were fixed and immunostained with rabbit anti-CXCR4 antibody (Abcam) at 1:250 for 12 hours. Cells were then incubated with Alexa 488 donkey anti-rabbit polyclonal antibody (Invitrogen) at 1:250 for 1 hour at room temperature. The cells were counterstained with DAPI to visualize nuclei. Immunofluorescence images were acquired with Zeiss ZEN 2010 software and a Zeiss LSM 710 confocal laser fluorescence scanning microscope.

5.1.2.6 Animal Care and Biosafety

Male CD-1 mice weighing 25 to 30 g were obtained from Charles River. Food and water were provided ad libitum. This investigation conformed with the Guide for the Care and Use of Laboratory Animals published by the US National Institutes of Health (NIH publication No. 85-23, revised 1996) and was approved by the Institutional Animal Use and Care Committee of the University of Pennsylvania (protocol #709026).

5.1.2.7 Ischemic Cardiomyopathy Model

Mice were anesthetized with ketamine (100 mg/kg) and xylazine (10 mg/kg), intubated with a 22-gauge catheter, and mechanically ventilated (Hallowell EMC). With the animal supine, an anterior thoracotomy was performed in the left fourth intercostal space, and an 8-0 polypropylene suture was placed around the mid left anterior descending coronary artery (LAD) midway between the left atrial appendage and left ventricular (LV) apex and ligated to produce a large anterolateral myocardial infarction of 30% of the LV. The extent of infarction is highly reproducible in our hands, and progression to cardiomyopathy has been well documented. After ligation, animals were randomly assigned to receive direct intramyocardial injection into the peri-infarct border zone of either saline (30 μL, n=7), SDF (6 μg/kg in 30 μL, n=13), or ESA (6 μg/kg in 30 μL, n=10). The thoracotomy was then closed, and the animals were extubated and recovered. Buprenorphine (0.5 mg/kg) was administered for postoperative analgesia. Both treatment groups received subcutaneous injections of 40 μg/kg liquid sargramostim (granulocyte-macrophage colony-stimulating factor), diluted in saline for a total volume of 100 μL, immediately postoperatively and on postoperative day 1.

5.1.2.8 In Vivo Angiogenic Growth Factor Expression

To assess upregulation of angiogenic growth factors and evaluate any potential sustained angiogenic effect of ESA treatment, hearts from all animals (control n=7, SDF n=13, and ESA n=10) were explanted 2 weeks after LAD ligation, and myocardial tissue biopsy samples were taken from the ischemic border zone. Samples were homogenized in T-Per tissue extraction reagent (Thermo-Fischer), normalized for total protein content via a Quick Start Bradford Protein Assay (Bio-Rad Laboratories), and tested for presence of mouse angiopoietin-1. Immunoblotting was performed with a mouse monoclonal antibody directed against mouse angiopoietin-1 (1:250; Abcam) and horseradish peroxidase–conjugated sheep anti-mouse IgG enhanced chemiluminescence secondary antibody (1:20 000; GE Healthcare). Chemiluminescent SuperSignal West Dura Extended Duration Substrate (Thermo Scientific) was used, and images were captured with a ChemiDoc XRS+ system (Bio-Rad). Assays were performed in triplicate.

5.1.2.9 In Vivo Inflammatory Analysis

To assess upregulation of inflammatory factors, hearts from all animals (control n=7, SDF n=13, and ESA n=10) were explanted 2 weeks after LAD ligation, and myocardial tissue biopsy samples were taken from the ischemic border zone. Samples were homogenized in T-Per Tissue Extraction Reagent (Thermo-Fischer), normalized for total protein content via

Quick Start Bradford Protein Assay (Bio-Rad), and tested for the presence of mouse tumor necrosis factor-α. Immunoblotting was performed with a rabbit polyclonal antibody directed against mouse tumor necrosis factor-α (1:1000; Abcam) and horseradish peroxidase–conjugated sheep anti-rabbit IgG enhanced chemiluminescence secondary antibody (1:3000; GE Healthcare). Chemiluminescent SuperSignal West Dura Extended Duration Substrate (Thermo Scientific) was used, and images were captured with a ChemiDoc XRS+ system (Bio-Rad).

To further assess the inflammatory response after treatment, hearts from a subset of animals (control n=4, SDF n=4, and ESA n=4) were explanted 2 weeks after LAD ligation, and myocardial tissue biopsy samples were taken from the ischemic border zone. Samples were homogenized in T-Per Tissue Extraction Reagent (Thermo-Fischer), normalized for total protein content via Quick Start Bradford Protein Assay (Bio-Rad), and tested for the presence of multiple inflammatory markers, including monocyte chemotactic protein-1, stem cell factor, nuclear factor (NF)-κB, phospho-NF-κB, phospho-p38 mitogen-activated protein kinase, phospho-Stat3, and phospho-IκB-α. Levels were quantified with the mouse inflammation ELISA strip (Signosis) and the PathScan Inflammation Multi-Target Sandwich ELISA (Cell Signaling Technology).

5.1.2.10 Echocardiographic Assessment of LV Geometry and Function

LV geometry and function were evaluated at 2 weeks (control n=7, SDF n=13, and ESA n=10) after LAD ligation with a high-resolution (30 MHz) Vevo 770 transthoracic echocardiography system (VisualSonic Inc). Four equally spaced short-axis acquisitions and 1 maximum long-axis cine-loop acquisition of the LV were recorded with the ultrahigh-frame-rate EKVTM (ECG-based kilohertz visualization) acquisition mode. LV systolic function was evaluated by a modified 4-plane Simpson method. LV cross-sectional areas were obtained by tracing the diastolic and systolic endocardial borders in each of the 4 equally spaced LV short-axis slices. On the basis of these measurements and the Simpson length (d, s), the fractional area change, ejection fraction, stroke volume, and cardiac output were computed with the Vevo 770 Standard Measurement Package. Before and during the ultrasound scanning, the mouse was lightly anesthetized with a mixture of 1% to 2% isoflurane gas and 100% oxygen and was placed in the supine position on a heating platform. All analyses were performed by a single investigator in a group-blinded fashion.

5.1.2.11 Statistical Analysis

Overall comparisons between the 3 groups were analyzed with a 1-way ANOVA. The results are reported with F (a ratio of the variance between groups to the variance within groups), 2 degrees of freedom, and the probability value. In addition, the unpaired Student t test was used to compare individual groups. Values are expressed as mean±SEM. Statistical significance was defined by P≤0.05.

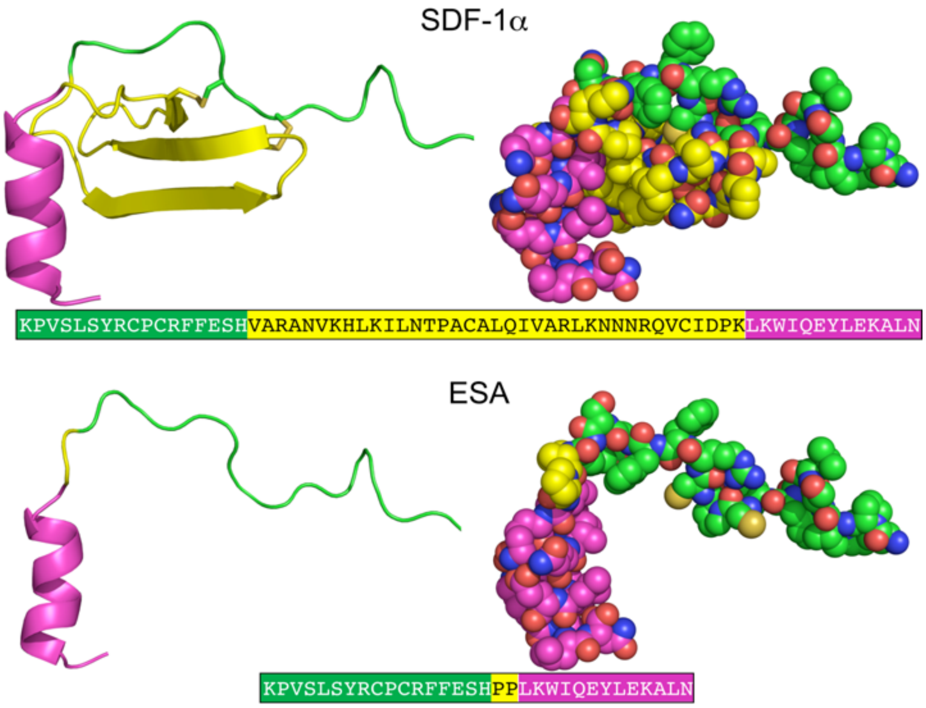


Figure 5.1. Crystallographic Structures of Stromal Cell–Derived Factor–1α (SDF) and Designed Model Structure ESA. The different regions of the structure are colored as N–terminal (green), central

region (yellow), and C–terminal (magenta). The central β-sheet region (yellow) is replaced by a diproline linker in ESA. The corresponding amino acid sequences of SDF–1α and ESA are also depicted, with the different regions colored accordingly.

5.1.3 Results

5.1.3.1 Computational Protein Design, Modeling, and Synthesis

The crystallographic structure of SDF at 2.2 Å resolution was used to guide design and modeling (Figure 5.1) [27]. The central β-sheet region residues (amino acids 18 to 54) were replaced with a 2-proline linker, which resulted in the 32-residue designed polypeptide ESA. The coordinates of residues 14FESHPPL20 were allowed to relax on modeling with MODELLER, whereas the coordinates of the remainder of the structure were constrained at their crystallographic values. Fifty different possible structures were generated for peptides with a 1-, 2-, or 3-proline linker. In general, the number of stable conformations increases substantially as the number of linking amino acids is increased (Figure 5.2). The peptide with the 2-proline linker (subsequently designated ESA) yielded energetic and conformational advantages. The ESA peptide structure model with the lowest-scoring function was chosen for further analysis. Importantly, the 2-proline residues were found to have energetically favorable backbone dihedral angles (ϕ and ψ) consistent with those observed for proline in natural protein structures. The introduction of a 2-consecutive-proline linker does not produce steric clashes within the modeled polypeptide structure. In addition, based on the calculated structure, phenylalanine 14 ($F14^{ESA}$) interacts with glutamic acid 15 ($E15^{ESA}$), histidine 17 ($H17^{ESA}$), and leucine 20 ($L20^{ESA}$ or L55 in SDF). These interactions are not present in the crystallographic structure of SDF, and this group of residues may form a small clusterlike structure that when coupled with the diproline spacer biases the peptide toward conformations similar to those found in native SDF (Figure 5.3). Further structural analysis revealed that the model did not contain any steric clashes or unusual conformations of the backbone and the amino acid side chains.

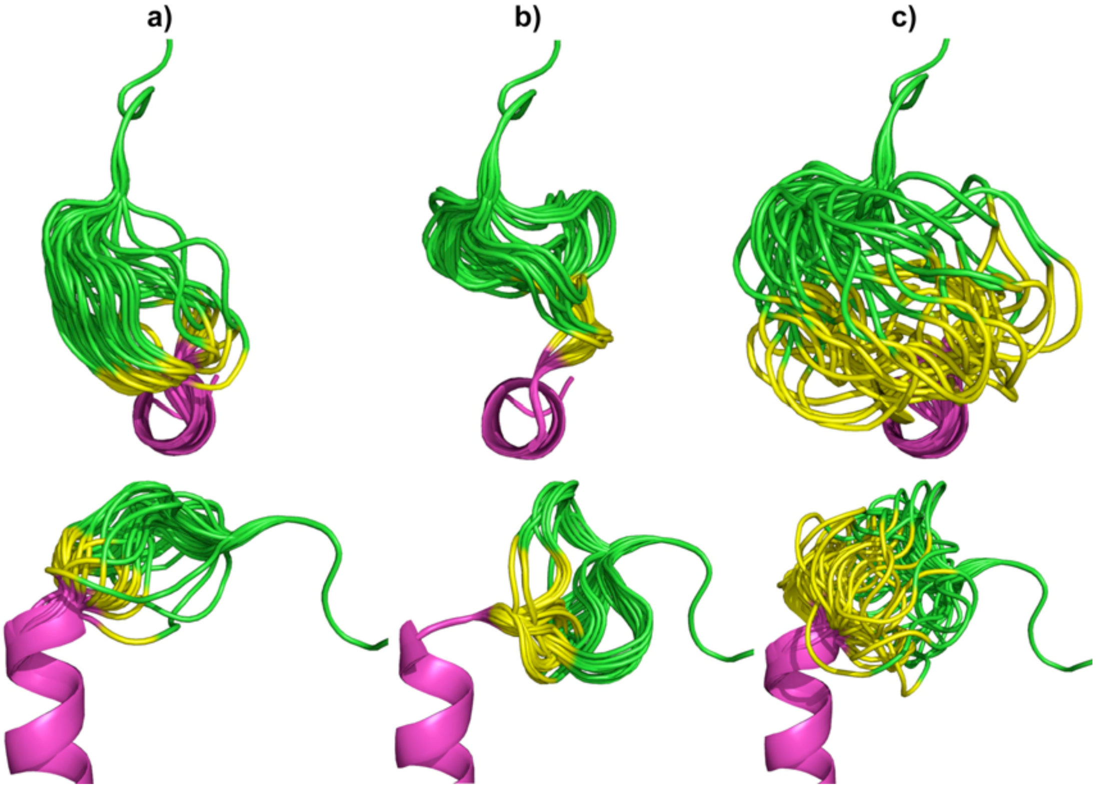


Figure 5.2. Top and Side View of Model Stromal Cell–Derived Factor–1α (SDF) Analog Peptides Using 1-proline (a), 2-proline (b), and 3-proline (c) Residues to Link the N and C Terminus. The images depict a composite of the 50 most energetically stable linker conformations of each peptide sequence. The peptide with the 2-proline linker (**b**) adopts a more uniform tertiary profile than the others and recovers the perpendicular orientation between the N– and C–termini found in native SDF.

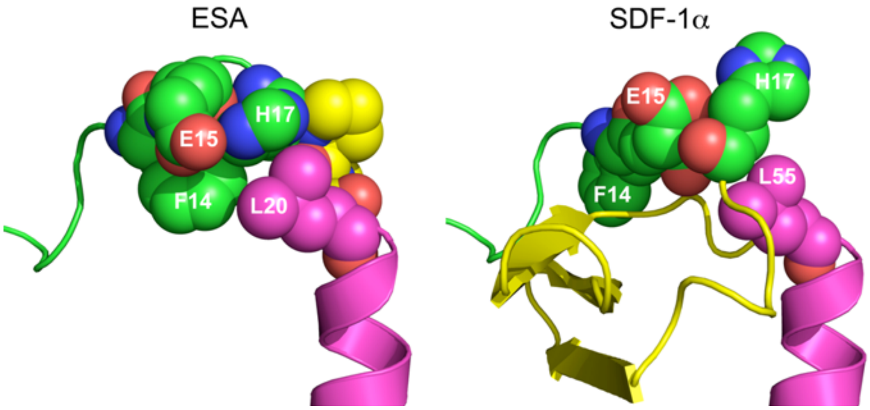


Figure 5.3. Space-Filling Representation of the ESA and SDF-1α Structures. The bulky side chain of phenylalanine 14 (F14^{ESA}) interacts with glutamic acid 15 (E15^{ESA}), histidine 17 (H17^{ESA}), and to some extent leucine 20 (L20^{ESA} or residue 55 in native SDF-1α). The interactions of F14^{SDF}, E15^{SDF}, H17^{SDF}, and L55^{SDF} were absent in the crystallographic structure of SDF–1α. These new interactions formed a small clusterlike structure that when coupled with the diproline spacer (yellow in ESA) may help to provide the necessary conformational stability and rigidity found in native SDF.

5.1.3.2 EPCs Cultured in Endothelial Specific Media Exhibit Enhanced Migration When Exposed to an ESA Gradient

Using a Boyden chamber assay, 14-day cells showed increased migration when placed in an ESA gradient compared with both recombinant SDF and saline (ESA 567±74 cells/high-power field [HPF] versus SDF 438±46 cells/HPF [*P*=0.037] versus control 156±45 cells/HPF [*P*=0.001]). Representative slides are shown in Figure 5.4. In addition, a 1-way ANOVA was used to generate an overall comparison of the 3 study groups. EPC migration differed significantly across the 3 groups, *F*(2,6)=41.27, *P*=0.0003. A dose-response curve generated by comparing molar equivalents of ESA and SDF (Figure 5.5) demonstrated that ESA significantly outperformed SDF at the high end of the concentration range but was essentially equivalent to SDF for the low- and mid-range concentrations. Additionally, it was revealed that the effectiveness of ESA and SDF was not linearly correlated to the dose concentration and that there was a significant decrease in EPC migration at the highest doses for both.

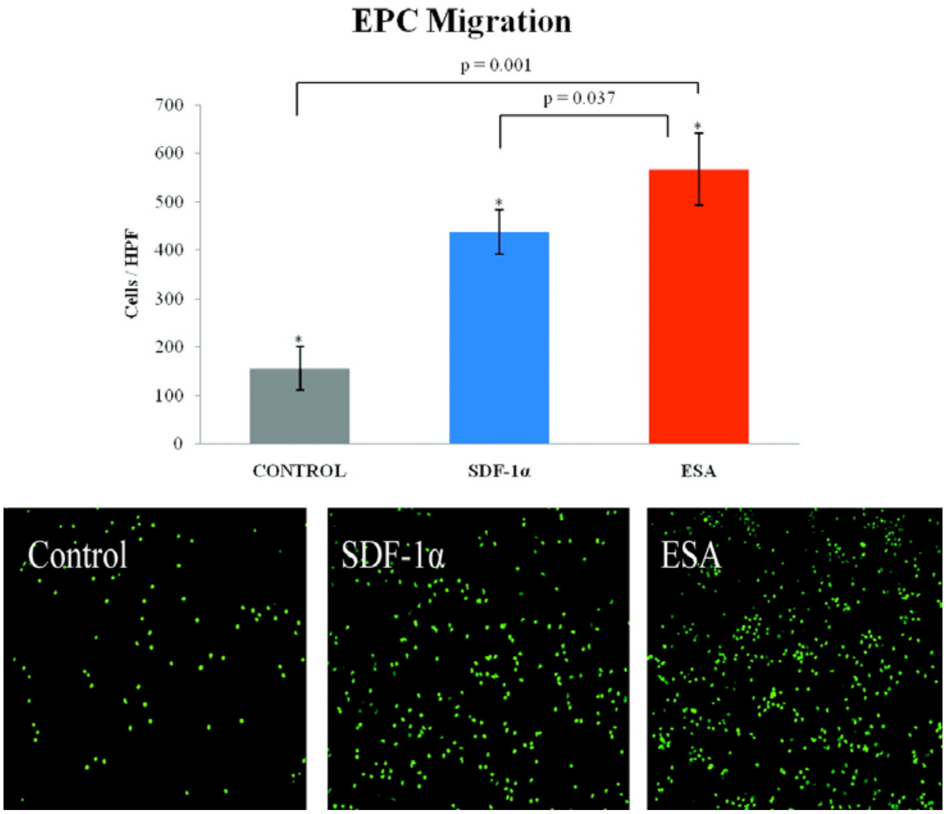


Figure 5.4. Boyden Chamber Assays. Representative images of Boyden chamber assay filters demonstrating green fluorescent protein–positive endothelial progenitor cell (EPC) migration when exposed to a saline, stromal cell–derived factor–1α (SDF), and ESA gradient, respectively. HPF indicates high-power field.

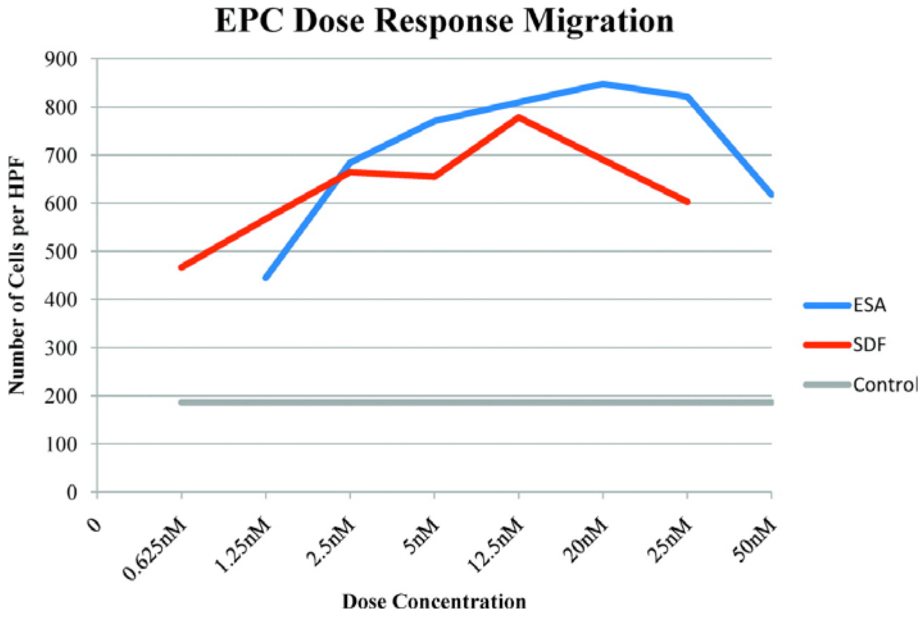


Figure 5.5. EPC Dose Response Migration. Dose-response curve of endothelial progenitor cell (EPC) migration comparing molar equivalent concentration gradients of ESA and stromal cell–derived factor–1α (SDF). HPF indicates high-power field.

5.1.3.3 ESA Elicits Increased CXCR4 Receptor Activation In Vitro

Cells treated with ESA had greater levels of AKT phosphorylation than either recombinant SDF (1.638±0.239 versus 1.258±0.187 OD, *P*=0.006) or control (1.638±0.239 versus 0.949±0.077 OD, *P*=0.0003). There was no significant difference in total AKT content among the groups. In addition, a 1-way ANOVA was used to generate an overall comparison of the 3 study groups. Levels of AKT phosphorylation differed significantly among the 3 groups, *F*(2,21)=38.56, $P<1\times10^{-7}$. The addition of the CXCR4 receptor antagonist AMD3100 to the culture media eliminated any difference in receptor activation between the groups (ESA 0.598±0.22 versus SDF 0.599±0.07 OD, P=0.49, and ESA 0.598±0.22 versus control 0.539±0.05 OD, *P*=0.27).

5.1.3.4 EPCs Display Robust CXCR4 Expression When Treated With ESA

After exposure to ESA or SDF for 6 hours, isolated EPCs demonstrated diffuse expression of the CXCR4 receptor (Figure 5.6).

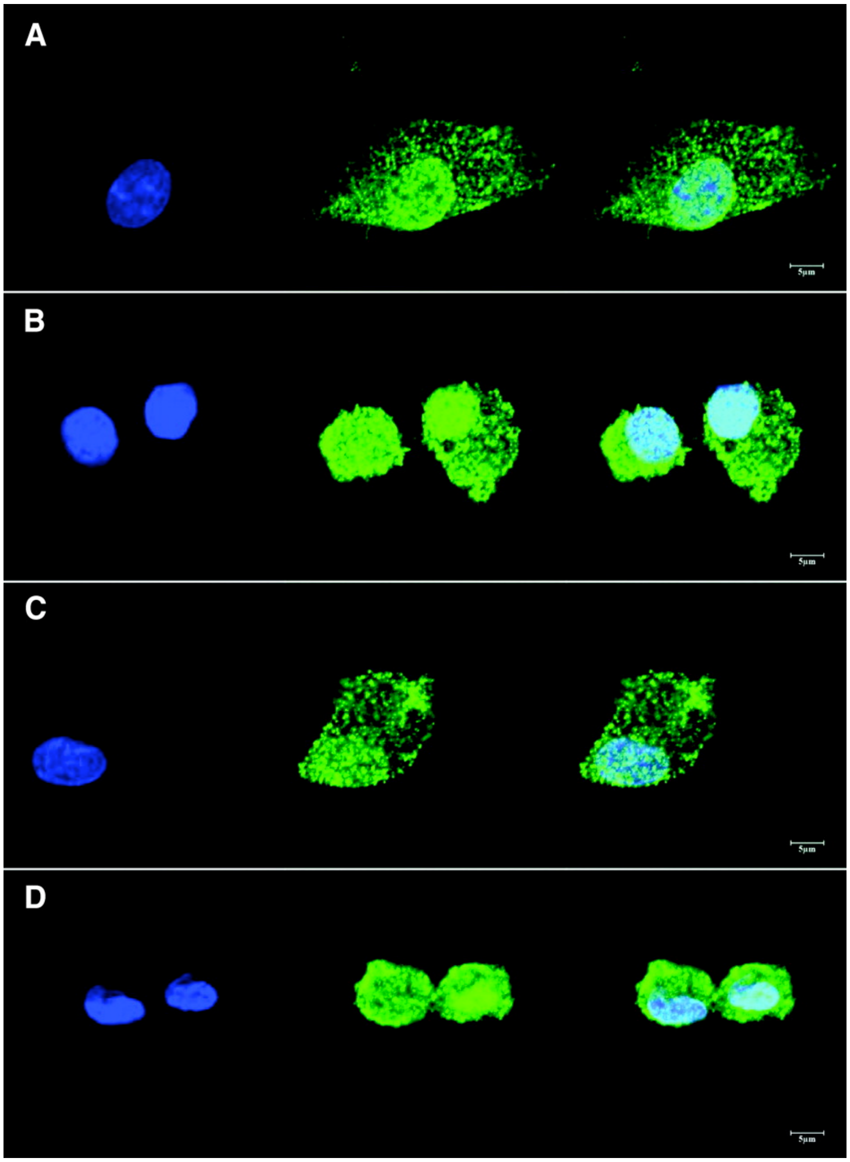


Figure 5.6. CXCR4 Expression in EPCs. Isolated endothelial progenitor cells cultured and stimulated with either stromal cell–derived factor–1α (A and B) or ESA (C and D) demonstrated robust and universal expression of the chemokine receptor 4 (CXCR4 receptor; green). Nuclei are stained with DAPI (blue).

5.1.3.5 In Vivo Angiogenic Growth Factors Are Upregulated by ESA

At 2 weeks, immunoblotting revealed a significant quantitative increase in mouse angiopoietin-1 expression in the ESA- and SDF-treated animals compared with the control group. Quantitative analysis demonstrated intensities of 336±9 (ESA), 367±17 (SDF), and 234±7 (control) intensity units, respectively ($P<0.001$). A representative immunoblot is shown in Figure 5.7.

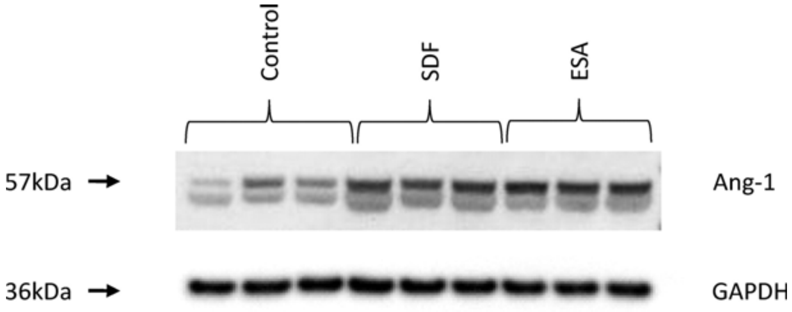


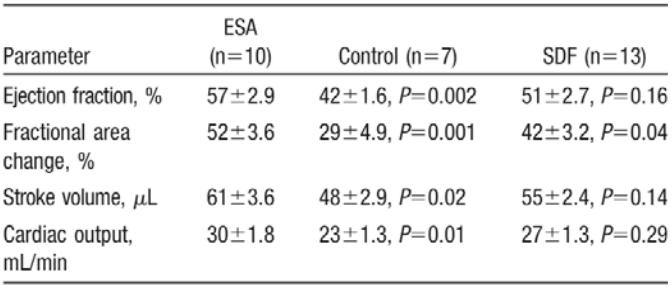
Figure 5.7. In vivo Mouse Angiopoietin-1 Growth Factor. At 2 weeks, border-zone myocardial tissue samples show increased levels of mouse angiopoietin-1 (Ang-1) in the ESA-treated and stromal cell–derived factor–1α (SDF)-treated groups compared with the control group. GAPDH staining was performed to demonstrate equivalent protein loading between lanes.

5.1.3.6 In Vivo Levels of Border-Zone Inflamatory Markers Are Not Statistically Different Between Groups

At 2 weeks, immunoblotting revealed similar levels of tumor necrosis factor-α among the ESA, SDF, and control groups. Quantitative analysis demonstrated intensities of 4784±274 (ESA), 4832±547 (SDF), and 4832±722 (control) arbitrary intensity units, respectively. Statistical comparisons revealed no statistically significant difference (ESA versus control, $P=0.46$; ESA versus SDF, $P=0.4$, SDF versus control, $P=0.5$). Additional analysis demonstrated that after infarction and treatment, border-zone levels of the inflammatory factors monocyte chemotactic protein-1, stem cell factor, NF-κB, phospho-NF-κB, phospho-p38 mitogen-activated protein kinase, phospho-Stat3, and phospho-IκB-α were either not significantly different or were significantly less in animals treated with ESA or SDF than in saline control animals. These findings indicate that inflammation is not responsible for the beneficial and proangiogenic effects of ESA.

5.1.3.7 Echocardiographic Assessment Demonstrates Enhanced LV Function After ESA Treatment

Echocardiographic assessment of cardiac function demonstrated significant benefits in the ESA group compared with the control group. At 2 weeks, ESA animals had a significantly improved ejection fraction (57±2.9% versus 42±1.6%, *P*=0.002), cardiac output (30±1.8 versus 23±1.3 mL/min, *P*=0.01), stroke volume (61±3.6 versus 48±2.9 μL, *P*=0.02), and fractional area change (52±3.6% versus 29±4.9%, P=0.001) compared with controls. The ESA-treated mice also had a significantly increased fractional area change compared with SDF–treated mice (52±3.6% versus 42±3.2%, *P*=0.04) and a strong trend toward improvement in all other ventricular function parameters assessed (Table 5.1). A 1-way ANOVA was used to generate an overall comparison of the 3 study groups for each cardiac functional parameter. Values differed significantly across the 3 groups for ejection fraction (*F*(2,27)=5.71, *P*=0.009), cardiac output (*F*(2,27)=3.98, *P*=0.03), stroke volume (*F*(2,27)=3.83, *P*=0.03), and fractional area change (*F*(2,27)=7.91, *P*=0.002).



Parameter	ESA (n=10)	Control (n=7)	SDF (n=13)
Ejection fraction, %	57±2.9	42±1.6, *P*=0.002	51±2.7, *P*=0.16
Fractional area change, %	52±3.6	29±4.9, *P*=0.001	42±3.2, *P*=0.04
Stroke volume, μL	61±3.6	48±2.9, *P*=0.02	55±2.4, *P*=0.14
Cardiac output, mL/min	30±1.8	23±1.3, *P*=0.01	27±1.3, *P*=0.29

Table 5.1. Left Ventricular Function Assessed by Transthoracic Echocardiography. At 2 weeks, ESA-treated animals demonstrated improved left ventricular function by transthoracic echocardiography. SDF indicates stromal cell–derived factor–1α.

5.1.4 Discussion

Angiogenic cytokine therapy for ischemic heart disease has proven to have great potential in numerous preclinical and clinical trials. The ability to replenish myocardial microvasculature could prove lifesaving to the millions of patients with myocardial ischemia. SDF, in particular, is among the most potent and specific angiogenic cytokines; its sole target, the CXCR4 cell surface antigen, is expressed in significant levels on $CD34^+$ EPCs, and expression of this receptor is related to efficient SDF-induced transendothelial migration [28]. SDF repeatedly has been shown to play a critical role in the rescue of myocardial function and stem cell recruitment to the heart after myocardial infarction [19,20,29-33]. However, significant barriers remain between the diverse experimental successes of SDF angiogenic cytokine therapy and its widespread clinical translation. SDF is a large, complex 10-kDa protein that is expensive to manufacture and is readily inactivated by proteases that are upregulated at the time of myocardial infarction [34,35]. It is our belief that smaller, synthetic analogs of SDF can provide translational advantages over the native protein, including enhanced stability and function, ease of synthesis, lower cost, and potential modulated delivery via engineered biomaterials. In the present study, we used advanced computational protein design techniques to create a more efficient and translatable molecule (ESA) evolved from the native SDF protein.

The ESA peptide was engineered to link the native N–terminus (responsible for receptor activation and binding) and the C–terminus (responsible for extracellular stabilization) with a diproline amino acid spacer. Although ESA may not have a well-defined tertiary structure, the linker is designed to bias the polypeptide toward structures that present the N– and C–termini in a manner similar to that present in the native SDF protein. On the basis of the modeled structure, $F14^{ESA}$ is in proximity to $E15^{ESA}$, $H17^{ESA}$, and $L20^{ESA}$, forming a small cluster of side chains within ESA (Figure 5.3). These interactions, coupled with the diproline linker, may provide a conformational bias sufficient to recover the activity associated with

SDF. The end product, ESA, is a novel polypeptide that contains less than half the number of amino acid residues, is less than half the molecular weight, and has enhanced physiological performance compared with recombinant SDF.

The SDF chemotactic homing mechanism is central to its ability to increase peri-infarct microvasculature and prevent mechanically inefficient ventricular contraction and eventual heart failure. Using a Boyden chamber assay to directly quantify the magnitude of cellular migration, we were able to demonstrate significantly enhanced migration of endothelial progenitor stem cells when placed in an ESA gradient compared with native SDF peptide. This result shows that for the same mass concentration, ESA not only retains this important function, despite its reengineered conformation, but surpasses the native protein. Compared with SDF, the relatively small size of ESA may enhance its diffusion potential and the speed at which the chemotactic signal is deployed. However, rapid diffusion is likely not the only factor that leads to significantly greater EPC migration. It has been shown that AKT activation is required for SDF-induced cellular migration [36]. To evaluate the EPC response to ESA, we incubated EPCs with culture media that contained ESA, recombinant SDF, or media only (control) and quantified both phosphorylated and total AKT levels by ELISA. In concordance with the results of the EPC migration assay, we were able to demonstrate significantly increased CXCR4 receptor activation and phosphorylated AKT levels in response to ESA compared with SDF. The addition of the CXCR4 antagonist AMD3100 eliminated this difference in receptor activation between ESA, SDF, and control, which confirms its receptor specificity. It is difficult to speculate which specific receptor-peptide interactions are responsible for the enhanced activation of CXCR4 by ESA, and future computational studies will be used to understand the surprisingly efficient activity of this novel polypeptide on a molecular level.

It is our belief that ESA, in a similar fashion to SDF, increases border-zone microvasculature, which in turn reverses cellular ischemia, preserves cardiomyocyte viability, and increases the mechanical efficiency of peri-infarct myocardium. In support of that hypothesis, we have shown that border-zone myocardium treated with ESA has significantly upregulated levels of angiopoietin-1, indicative of an ongoing angiogenic process as late as 2 weeks after infarction and treatment. One could argue that the mechanism responsible for the upregulated angiogenic activity is an exaggerated inflammatory response that results from the injection of ESA or even SDF. However, analysis of border-zone tumor necrosis factor-α, monocyte chemotactic protein-1, stem cell factor, NF-κB, phospho-NF-κB, phospho-p38 mitogen-activated protein kinase, phospho-Stat3, and phospho-IκB-α revealed that levels of inflammatory markers were not increased in experimental animals compared with controls, which indicates that the beneficial angiogenic and functional effects were not the result of differing levels of inflammation.

An increase in microvascular perfusion should result in decreased ventricular remodeling, improved regional ventricular function, and slower progression toward heart failure. Left ventricular geometry and function were evaluated at 2 weeks after LAD ligation with a high-resolution (30 MHz) transthoracic echocardiography system. The improved functionality of ESA in vitro was, once again, borne out in vivo. Animals treated with intramyocardial injections of the ESA peptide had better ejection fractions, cardiac output, stroke volume, and fractional area contraction than either the SDF or control group.

5.1.5 Conclusion

In conclusion, we have been able to engineer a novel, low-molecular-weight polypeptide with the enhanced physiological ability to induce EPC migration and EPC receptor activation and improve ventricular performance compared with native SDF. This peptide offers a more clinically translatable neovasculogenic therapy that could conceivably be deployed at any point in the time course of ischemic heart disease to address critical deficits in microvascular perfusion.

5.2 Molecular Dynamics Simulations of Stromal Cell–Derived Factor–1α and Simplified Peptide Analogs with Enhanced Chemotactic Activity[1]

5.2.1 Introduction

Chemokines are a family of small (8−10 kDa) secreted proteins that posses chemoattractant properties [37,38]. This type of cytokines directs cell migration and proinflamatory immune responses by specific activation of their receptors expressed on the cell surface. The chemokine receptors are part of the large family of seven−transmembrane G protein−coupled receptors (GPCRs) [39]. In addition, chemokines play an important role in promoting the creation and regulation of new blood vessels (angiogenesis and angiostasis) [37,40]. Chemokines are divided in four families based on the arrangement and number of cysteine residues in the N−terminus, i.e., CXC, CC, CX_3C, and C [37,38].

Stromal cell−derived factor−1 (SDF−1 or CXCL12) is a member of the CXC chemokine family and is responsible for the migration of a variety of cells, including monocytes and lymphocytes [38]. It is constitutively secreted from the bone marrow stromal cells and it plays a central role in the hematopoietic stem cell trafficking [41,42]. SDF−1 is the ligand of the CXC chemokine receptor 4 (CXCR4). The SDF-1/CXCR4 signal axis has been shown to play essential roles in: hemotopoiesis, stem/progenitor cell trafficking, stem-cell homing, cancer progression-growth, metastasis, neovascularization and angiogenic processes [11,43-46]. Furthermore, mice lacking the SDF–1 or the CXCR4 gene die in utero with a number of developmental abnormalities indicating the critical role of SDF–1 and CXCR4 in embryonic development [47-49].

Three variants of SDF−1 are produced by alternative splicing of the same gene, i.e., SDF–1α, SDF–1β, and SDF–1γ [50]. SDF-1α is the most abundant form and is composed of 68 residues. SDF–1β contains four additional residues in the C−terminus while the SDF–1γ variant contains a more extensive C−terminus with a different amino acid composition. SDF–

1α assumes the conventional chemokine topology as shown by its NMR [22,51] and crystal structures [27,52,53]. It is composed by an unstructured N–terminus (K1 to R8), a long flexible loop that contains the CXC conserved motif (denominated N–loop), a three-stranded anti-parallel β-sheet, and an α-helix. Interestingly, two disulfide bonds formed by residues at the N–loop and the β-sheet region, restrict the conformations of SDF–1α.

The interaction of SDF–1α with the CXCR4 receptor has been widely studied [38,45,46] and the current binding model suggests a two-step/two-site mechanism [22,54]. First is the interaction between SDF–1α (through its β-sheet, 50–s loop and N–loop) and the CXCR4 extracellular region. After SDF–1α is effectively anchored on the extracellular region of CXCR4, its highly dynamic N−terminus efficiently searches and binds a cavity within the transmembrane helices that triggers the conformational changes needed for signal transduction [54], pressumably by the fly-casting mechanism [55].

Based on the relevance of the SDF–1α/CXCR4 system, its potential role as a therapeutic target for treatment of different disorders (e.g., cancer progression-growth, metastasis, inflammatory and infectious diseases, etc) has been extensively explored [40,56]. Of particular relevance are its roles in the growth of new blood vessels from existing ones (angiogenesis) and the de novo formation of vessels from endothelial progenitor cells (vasculogenesis). These provide a promising treatment for inadequate blood supply to the heart (ischemic heart disease), which has been predicted to be the leading cause of death worldwide in the near future [11,46].

For instance, externally administered SDF–1α to ischemic myocardium has been demostrated to produce beneficial effects such as increase circulating endothelial progenitor cells (EPC), increment in the de novo formation of vessels (vasculogenesis), increase in myocardial function, and preservation of ventricular geometry, just to mention some [19,20].

Despite its potential as a therapeutic agent for treatment of cardiovascular disorders, SDF–1α presents features that preclude its translation to general clinical therapy.

Recombinant SDF–1α suffers proteolysis by different proteases [21,35], has a complex tertiary structure that involved two disulfide bonds [22], and is relatively large and expensive.

Hiesinger et al. recently proposed a promising alternative based on computationally designed SDF–1α analogs that used a proline residue linker to connect the N–terminus and part of the N–loop (K1 to H17) with the α-helix at the C–terminus (L55 to N67) [57]. The study showed that proline analogs performed similarly and in some cases, outperform recombinant SDF–1α in cell-based assays. Moreover, one of the analogs outperformed recombinant SDF–1α in a murine model of ischemic heart failure [57]. In this study, a detailed analysis of the conformations of SDF–1α and some simplified analogs was carried out. Molecular dynamics (MD) simulations were performed to obtain insights of the conformations explore by SDF–1α and the different analogs in water environment and under physiological conditions of temperature and pressure. MD simulations suggest that the analogs explore similar conformations as native SDF–1α, especially the two-proline (2P) analog. MD simulation results are in close agreement with experimental results using a migration assay and a receptor activation assay [57]. Hypotheses about the reasons of enhanced activity of the two-proline analog relative to native SDF–1α in both, cell-based assays and in the murine model of heart failure are exposed and discussed in regard to experimental findings.

5.2.2 Methods

5.2.2.1 Initial Structures

The crystal structure of SDF–1α at 2.2 Å resolution was used [27]. The complete structure was reconstructed using the CHARMM topology files. The modeling of the simplified SDF–1α analogs is described previously [57]. Briefly, the crystal structure of SDF–1α was used to guide the design. The segment of the protein composed by the β-sheet region was replaced by linkers made of one (1P), two (2P), and three (3P) proline residues using MODELLER [25]. An additional linker composed of two glycine residues was also considered

(2G). Finally, for the analog with a linker of three prolines, two mutations in the α-helix were suggested (I58E and L62K) to increase the helix stability (3P_I24E_L28K). The initial structures utilized for the molecular dynamics simulations are shown in figure 5.8.

5.2.2.2 Molecular Dynamics Simulations

All-atom molecular dynamic simulations were performed on the SDF–1α structure and the simplified analogs. The simulations were carried out using NAMD2 [58] and the CHARMM22 force field [59]. The best structure from the comparative modeling was minimized after addition of hydrogen atoms for 5000 steps using NAMD2 [58]. Water molecules (TIP3P) [60] and sodium (Na^+) and chlorine (Cl^-) ions [59] were included and minimized with the protein atoms fixed followed by a short MD simulation (5.0 ps with 1.0 fs time step). The protein was then minimized to adjust to the new environment for 10000 steps followed by another 10000 steps minimization of the entire system. Finally, the MD simulation on the entire system was carried out in the NPT ensemble. Langevin dynamics and the hybrid Nosé-Hoover Langevin piston were used to maintain constant temperature (310 K) and constant pressure (1atm), respectively [61]. Electrostatic interactions were evaluated using the Particle Mesh Ewald technique with a grid spacing less than 1.0 Å for each dimension and a fourth-order interpolation [62]. Periodic boundary conditions were used with a periodic cell of 55 Å × 52 Å × 69 Å. Bond lenghts involving hydrogen atoms were constrained to their equilibrium values [59] by using the SHAKE algorithm [63]. A time step of 2.0 fs was used during the simulations.

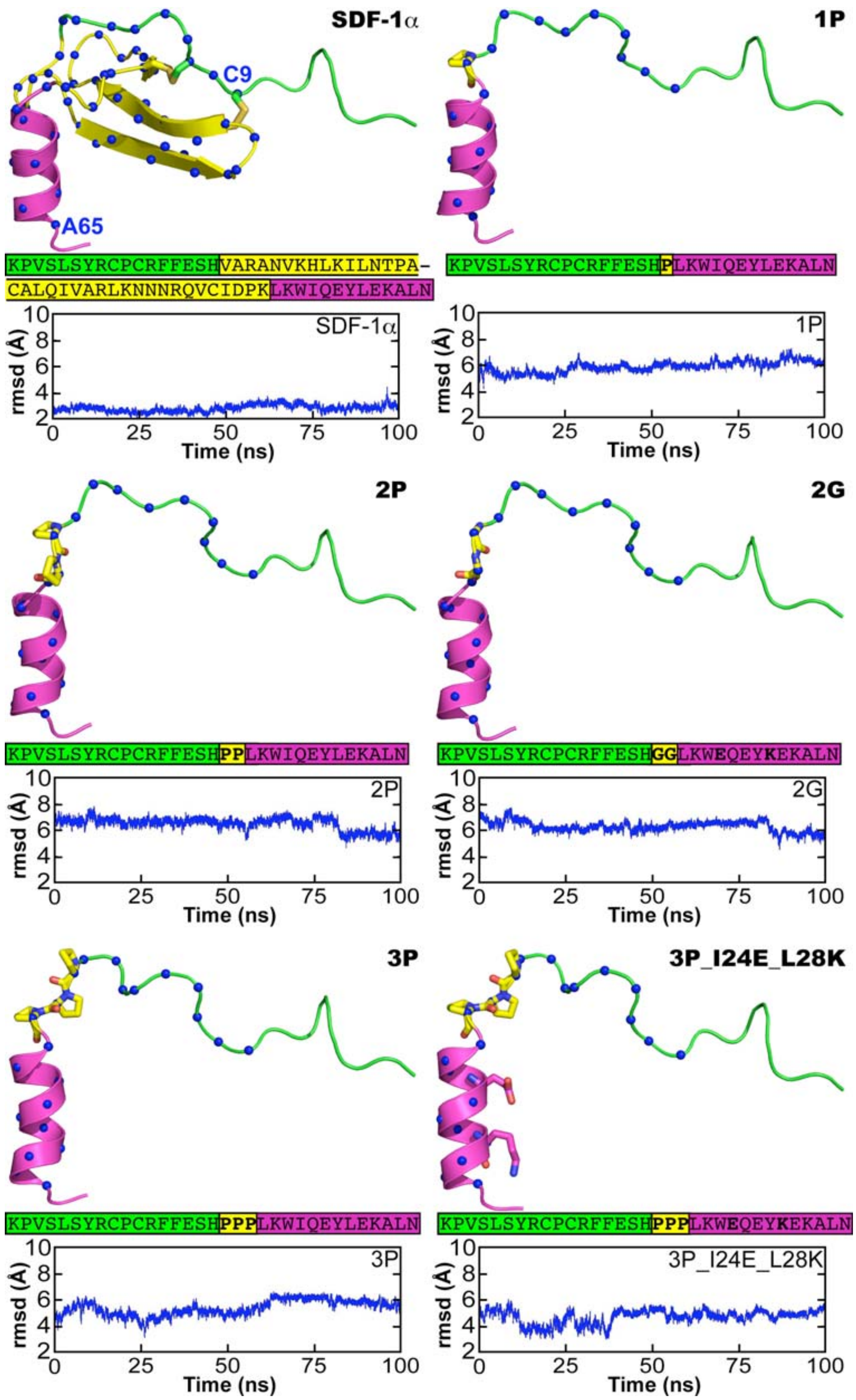

SDF-1α
C9
A65
KPVSLSYRCPCRFFESHVARANVKHLKILNTPA–
CALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
SDF-1α
rmsd (Å)
Time (ns)
1P
KPVSLSYRCPCRFFESHPLKWIQEYLEKALN
1P
rmsd (Å)
Time (ns)
2P
KPVSLSYRCPCRFFESHPPLKWIQEYLEKALN
2P
rmsd (Å)
Time (ns)
2G
KPVSLSYRCPCRFFESHGGLKWEQEYKEKALN
2G
rmsd (Å)
Time (ns)
3P
KPVSLSYRCPCRFFESHPPPLKWIQEYLEKALN
3P
rmsd (Å)
Time (ns)
3P_I24E_L28K
KPVSLSYRCPCRFFESHPPPLKWEQEYKEKALN
3P_I24E_L28K
rmsd (Å)
Time (ns)
2
4
6
8
10
0
25
50
75
100

Figure 5.8. Initial Structures (With Sequences) of SDF–1α and the Simplified Analogs Used for the Molecular Dynamics Simulations. The regions of SDF–1α and analogs are colored as follows: N–terminus in green, β-region/Linker in yellow, and C–terminus (including the α-helix) in magenta. The root-mean square deviation (rmsd) plots relative to the initial structure for the production phase (100 ns) are also shown. The Cα atoms involved in the calculation of the rmsd are depicted as blue spheres. To avoid noise from the highly dynamic N–terminus, the first eight residues were excluded in the rmsd calculations.

5.2.2.3 Overlap Distribution Calculation

The overlap distribution (Θ) calculation was done using the following equation:

$$\Theta(\nu_1,\nu_2) = \frac{1}{N}\left(\sum_{i=1}^{W}\frac{\nu_{1i}\cdot\nu_{2i}}{\sqrt{\nu_{1i}^2\cdot\nu_{2i}^2}}\right) \qquad (5\text{-}1)$$

where N is the total number of points, ν_{1i} and ν_{2i} are the values for the first and the second histograms, respectively. The sum is over the different "bins" in the histrograms. W is the number of bins. The denominator is the normalization factor. A value of 0 means there is not overlap for the two distributions whereas a value of 1 means a maximum ovelap (equal distributions).

5.2.3 Results

5.2.3.1 Structural Parameters and General Dynamics of the Systems

The equilibriation of the different systems was monitored by changes in structural parameters such as the root-mean square deviation (rmsd) relative to the initial structure. The production phase for each of the systems was 100 ns. In the case of SDF–1α, the rmsd of the Cα atoms was used to monitor the dynamics of the protein. It was evident from the results that the highly dynamic N–terminus was introducing noise in the analysis (figure 5.9a), so a subset of the 68 Cα atoms of SDF–1α was chosen (from C9 to A65). The rmsd of the Cα atoms from C9 to A65 is shown in figure 5.8 and 5.9a. For consistency, the equivalent residues in the SDF-1α analogs were used in the calculations of rmsd.

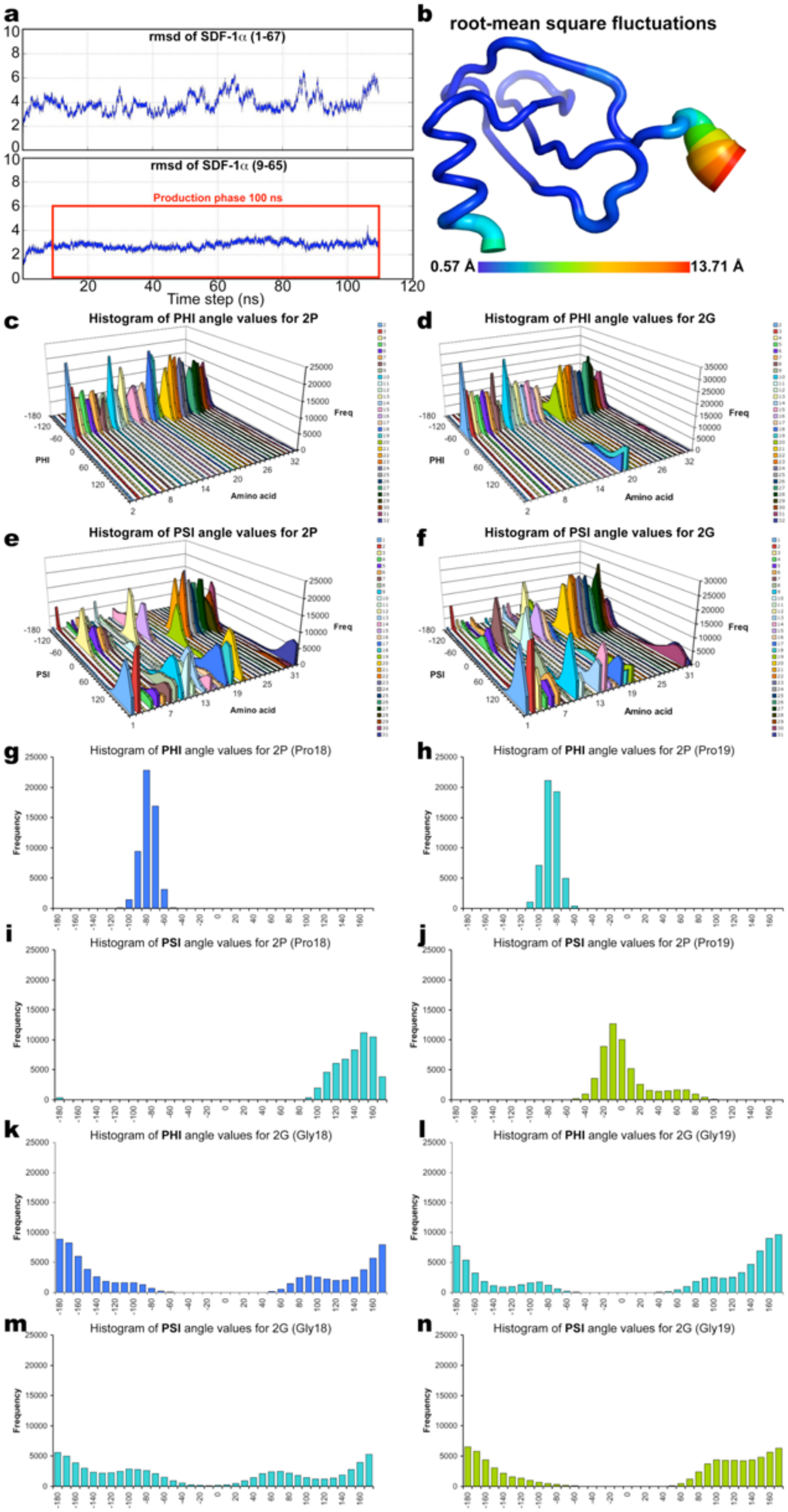

a
rmsd of SDF-1α (1-67)
rmsd of SDF-1α (9-65)
Production phase 100 ns
Time step (ns)
b
root-mean square fluctuations
0.57 Å
13.71 Å
c
Histogram of PHI angle values for 2P
d
Histogram of PHI angle values for 2G
e
Histogram of PSI angle values for 2P
f
Histogram of PSI angle values for 2G
PHI
PSI
Freq
Amino acid
g
Histogram of PHI angle values for 2P (Pro18)
h
Histogram of PHI angle values for 2P (Pro19)
i
Histogram of PSI angle values for 2P (Pro18)
j
Histogram of PSI angle values for 2P (Pro19)
k
Histogram of PHI angle values for 2G (Gly18)
l
Histogram of PHI angle values for 2G (Gly19)
m
Histogram of PSI angle values for 2G (Gly18)
n
Histogram of PSI angle values for 2G (Gly19)
Frequency

Figure 5.9. Root-mean Square Deviations (rmsd) and Root-mean Square Fluctuations (rmsf) of SDF–1α, and Distributions and PSI and PHI Backbone Angles of the 2P and 2G Analogs. (**a**) Rmsd for the entire trajectory (~110 ns) for SDF–1$\alpha_{1\text{-}67}$ and SDF–1$\alpha_{9\text{-}65}$. Inclusion of the highly dynamics N–terminus segment (K1-R8) produces large fluctuations in the rmsd that are not present when this segment is excluded from the calculations. The production phase is highlighted in red. (**b**) SDF–1α root-mean square fluctuations for the Cα atoms for the MD trajectory (100 ns). The thickness and color of the rederings are related with the rmsf value. (**c-f**) Distributions of the ϕ and ψ backbone angles for all the residues in the 2P and 2G analogs. (**g-n**) Distributions of ϕ and ψ angles for the residues that formed the linker in 2P and 2G analogs.

Additionally, the root-mean square fluctuations (rmsf) were calculated for the entire production phase and the results for SDF-1α are presented in figure 5.9b. From the rmsf results, it is evident the most dynamics part of the structure reside in the first segment of the N–terminus, K1 to R8. The disulfide bonds between C9-C34 and C11-C50, seem to significantly restrain the movement of the rest of the N–terminus (see figure 5.8).

In the case of the different analogs the conformations from the linker residues are in agreement with values seeing in natural proteins. As expected the proline residues forming the linker present a very narrow distribution of the phi (ϕ) and psi (ψ) angles through the entire trajectory while the glycine residues explore a relatively wide distribution, see Figure 5.9c-n.

5.2.3.2 Simplified Parameters to Monitor the Dynamics of SDF-1α and its Analogs

Instead of using techniques to reduce the dimentionality of the variables involved in the conformations explored by SDF–1α and the analogs, e.g., principal component analysis, we decided to identify simple parameters with straightforward physical interpretation. For this reason three parameters were defined and monitored: i) the distance between the Cα atoms of residue C9 and L55 (located before the beginning of the α-helix in the C–terminus), ii) the distance between the Cα atoms of residues C9 and A65 (last residue of the α-helix in the C–terminus), and iii) the angle formed by the Cα atoms of C9-L55 and A65, see figure 5.10.

To explore the similarities in the conformations sampled by SDF–1α and its simplified analogs, the distributions of the values adopted by the systems were obtained. The different histograms were compared and the overlaps calculated using equation 5-1. The results are presented in figure 5.10. The values of the overlap range from 0 (no distribution ovelap) and 1 (complete distribution overlap). Based on the distributions of these three parameters, it seems that the analogs sample similar conformations than SDF–1α. In the case of **distance1** (C9-L55), the 3P_I24E_L82K and 2G analog sample similar values than SDF-1α, but for the rest, there is still significant overlap in all the analogs but 1P. For **distance2** (C9-A65), the results shown a better overlap for 1P and 2P with significant values for all the rest analogs but 3P_I24E_L82K. In the case of **angle1** (C9-L55-A65), all the distributions seem to overlap significantly, with the oustanding case of the 1P and 2P analogs. It is our belief that the position of the N–terminus relative to the rest of the structure could be important in the interactions of SDF–1α with its native target receptors.

The findings from this analysis indicate that in general and based on the three parameters defined previously, all the simplified analogs of SDF–1α explore to a certain degree, conformations similar to the native SDF–1α. A product of the values for the overlap of the three parameters suggests that the rank of similarity between SDF-1α and the simplified analogs is 2P (0.083) > 2G (0.059) > 1P (0.015) > 3P (0.012) > 3P_I24E_L82K (0.003). The results based on this analysis are in agreement with experimental findings that show that the 2P analog bahaves similarly to native SDF–1α in a migration assay and a receptor activation assay [57].

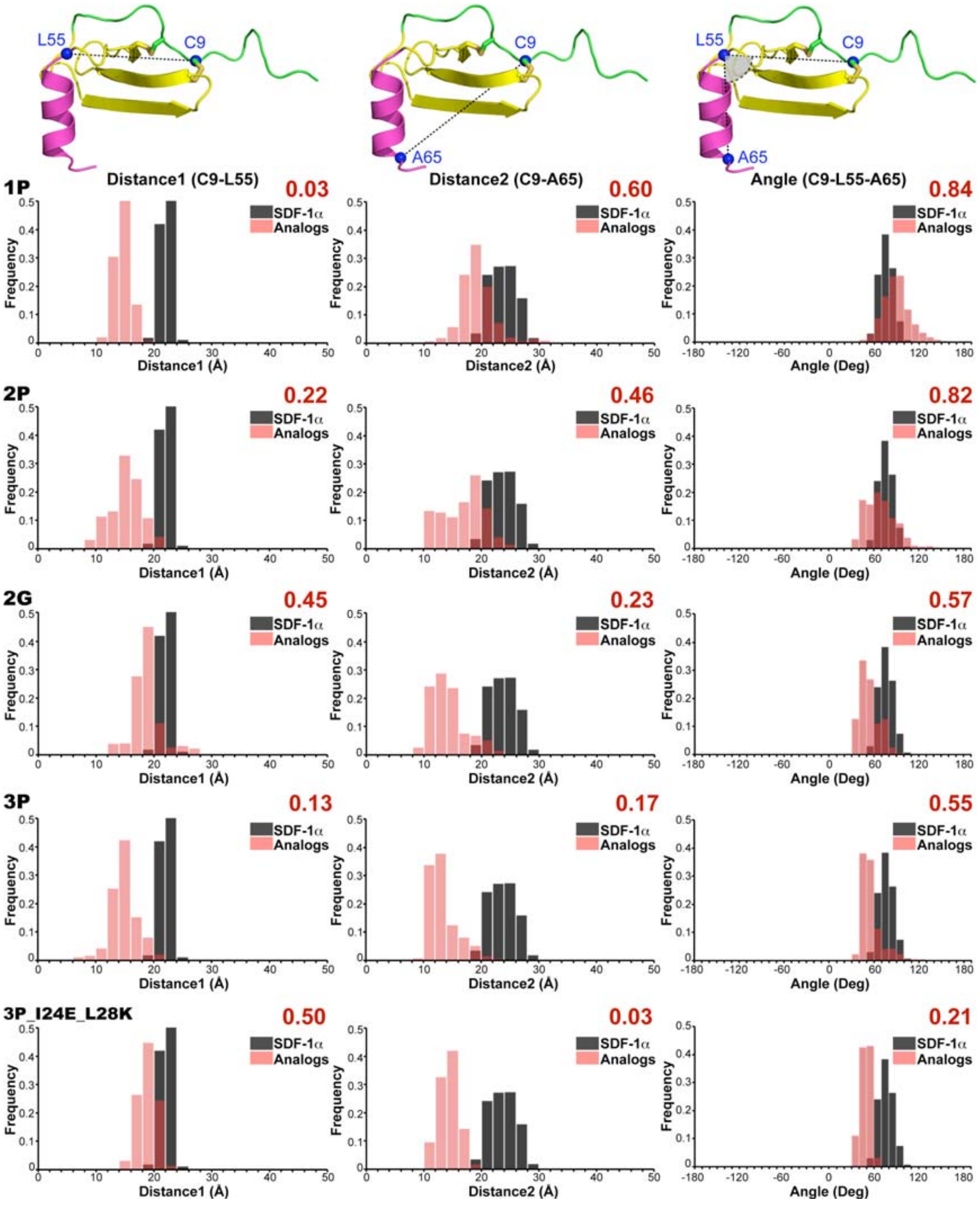


Figure 5.10. Parameter Defined to Monitor the Dynamics of SDF−1α and the Simplified Analogs. Three parameters, two distances and one angle are defined in the SDF−1α structure (top panels). The distributions of the values sampled along the MD simulations of the simplified analogs are compared with the distribution sampled by SDF−1α. The similarity (overlap) of the distributions is expressed by the

number at the top right corner of the histograms, were 0 represents no overlap and 1 represents a complete overlap.

5.2.3.3 Comparison of Superimposed Structures

A structure from the entire 100 ns MD trajectory was extracted every 2.0 ns and a comparison of the 50 resulting structures is presented in figure 5.11 for SDF–1α and the simplified analogs. From the MD trajectories different features for SDF–1α and each of the simplified analogs can be deduced.

SDF–1α: Based on the rmsd and the MD trajectories, the different conformations explored by SDF–1α at this time scale are similar ($\langle$rmsd$\rangle$ = 2.79 Å) than the initial structure (based on the X-ray crystal structure, 1A15.pdb; chain A [27]). As expected, the N–terminus is highly dynamic and explores a significant space volume with a radius of gyration for the nitrogen atom of the first residue (K1) of 14.79 Å. Starting at residue C9, the structure samples less diverse conformations that resemble the initial structure. As suggested previously, the presence of the two disulfide bonds (C9-C34 and C11-C50) seems to aid in maintaining the tertiary structure of SDF–1α. The secondary structure is well maintained during the simulation as indicated in figure 5.11 (Top left panel).

***1P** analog*: Interestingly, the MD simulations indicate a disruption of the α-helical structure in the analog with one-proline linker. The N–terminus is still very flexible with radius of gyration for the nitrogen atom of the first residue (K1) of 13.91 Å. The lack of the disulfide bonds seems to produce a highly flexible peptide, see figure 5.11, top right panel.

***2P** analog*: MD simulations indicate that even though there are no disulfide bonds present in the peptide, the general topology of 2P resembles the tertiary structure of native SDF–1α (figure 5.11, middle left panel). Contrary to the 1P analog, the secondary structure of the α-helix is retained. The N–terminus is highly flexible and explores an even larger space volume than SDF–1α as indicated by the radius of gyration of the nitrogen atom of the first residue, 17.13 Å.

***2G** analog*: The flexible residues in the linker of this peptide seem to allow the proximity of part of the N–terminus with the α-helix at the C–terminus (resembling an inverted "U" from the poses in figure 5.11, middle right panel). The structure of the α-helix is retained and similarly to SDF–1α, 1P and 2P, the N-terminus explores a significant space volume with a radius of gyration for the nitrogen atom of the first residue of 14.96 Å.

***3P** analog*: The MD simulations show a similarity in the general conformation explored by 2G and 3P. In a similar fashion, part of the N-terminus of 3P, samples conformations that are in the vicinity of the α-helix at the C-terminus, forming an inverted "U" topology, see figure 5.11, bottom left panel. The secondary structure of the α-helix is conserved and the N–terminus also explores a significant space volume with a radius of gyration of 13.10 Å for the nitrogen atom in the first residue.

***3P_I24E_L28K** analog*: Finally, a similar behavior than 2G and 3P is displayed regarding the proximity of the N– and C–terminus (inverted "U" topology, see figure 5.11 bottom right panel). Although the structure of the α-helix is conserved, the space volume explored by the N–terminus is dramatically reduced as suggested by the radius of gyration of the nitrogen atom of the first residue, 3.91 Å.

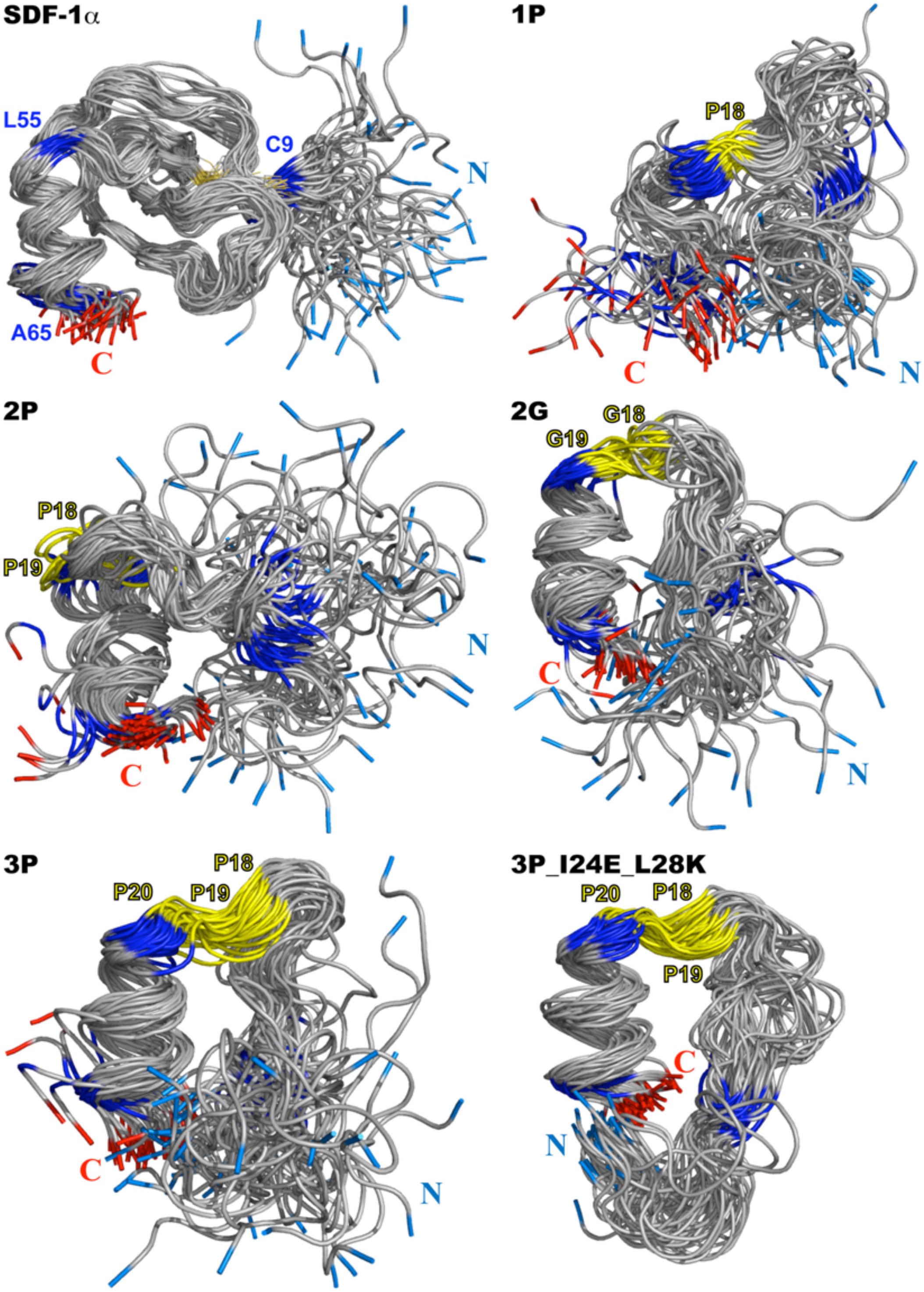


Figure 5.11. Superimposed Structures for SDF–1α and the Simplified Analogs Taken Every 2.0 ns for the Entire Production Phase (100 ns). For clarity and to indicate the positions of the N– and C–terminus, the first and last residues in the structures are colored in pale blue and red, respectively. In the

analogs, the residues forming the linker are colored in yellow. Finally, the positions that determined the parameters defined to monitor the dynamics of the structures are colored in blue.

5.2.3.4 Space Volume Sampled by the N–terminus, Radius of Gyration of the Center of Mass of Segment K1-R8

It has been shown that the N–terminus is responsible for the interaction and activation of SDF–1α with its native receptors [22]. To have further insights into the similarities of the conformations explored by the highly dynamic N–terminus of SDF–1α and the analogs, we simplified the degrees of freedom to only the center of mass of the segment K1-R8 and then compared the regions explored along the 100 ns of the MD trajectories.

Sampled structures along the MD simulations were aligned to the initial structure in each case. The alignment was based on the Cα atoms of the α-helix, e.g., K56 to A65 (or equivalent positions in analogs). Then, the center of mass of the Cα atoms of the segment K1-R8 was calculated (see small spheres in Figure 5.12). The cloud of points (and its center) generated for SDF-1α and the simplified analogs is shown in figure 5.12, in blue and red, respectively. In comparing the overlap, the initial structure of SDF–1α and the analogs were used and the transformation matrix needed to obtain the best-fit alignment between the analogs and SDF–1α was calculated. The atoms involved in the calculation were C9-H17 and K56-A65 (SDF-1α) and the equivalent residues in the analogs (the subset is composed of N–terminal residues involved in the rmsd calculation and the residues in the α-helix, i.e., residues that are common in all the structures). The same transformation matrix was applied to the points from the analogs and the overlap was calculated using equation 5-1. The values obtained for the overlap were: 1P (0.000), 2P (0.011), 2G (0.007), 3P (0.001), and 3P_I24E_L28K (0.000). Although small, the overlap for the center of mass of the N–terminus suggests that the 2P analogs displays more similarities (with SDF–1α) in the conformations explored than that rest of the analogs, see figure 5.12. Furthermore, the results from the overlap of the spatial region sampled by the N–terminus suggest the SDF–1α similarity rank:

2P > 2G > 3P > 1P = 3P_I24E_L28K. Based on the overlapping values it is not possible to discriminate between 1P and 3P_I24E_L82K, but based on the visual considerations, we could speculate that 3P_I24E_L82K would diverge more from SDF–1α in the conformation explored by both structures. The final similarity rank from this section is 2P > 2G > 3P > 1P > 3P_I24E_L28K. This is slightly different (order of 1P and 3P analogs) than the rank proposed previously (using three parameters), 2P > 2G > 1P > 3P > 3P_I24E_L82K.

In addition to the overlaps, figure 5.12 also contains information about the relative position of the center of the cloud of points for SDF–1α and the simplified analogs. In the case of SDF–1α, the center is located in front of the helical positions (see yellow spheres in figure 5.12) that face the β-region. The central region and the disulfide bonds seem to preclude the proximity of the N–terminus to the α-helix.

In the case of 1P, the center is located behind the "reference positions" (yellow spheres), which in the case of SDF–1α, face the β-region. Similar to SDF–1α, the center in the case of 2P is located in front of the reference positions that originally face the β-region. The absence of the β-region and the disulfide bonds allows the N–terminus to explore conformations closer to the α-helix. Similar to the center of 2P, the center of the 2G analog is located in front of the reference positions but perhaps due to the large flexibility of the residues composing the linker, the center is closer to the α-helix. The center of 3P is located in a very similar position as 2G. Finally, in the case of 3P_I24E_L28K, the center is very close to the α-helix located at the C–terminus.

From this section the results indicate the features of the different regions of space that are sampled by the N–terminus of SDF–1α and the simplified analogs. Again, the 2P analog seems to mimic better the conformations of native SDF–1α. In contrast, the conformations of the N–terminus of the 3P_I24E_L28K analog are restricted to a relative small volume and also they diverge more from those explored by SDF–1α.

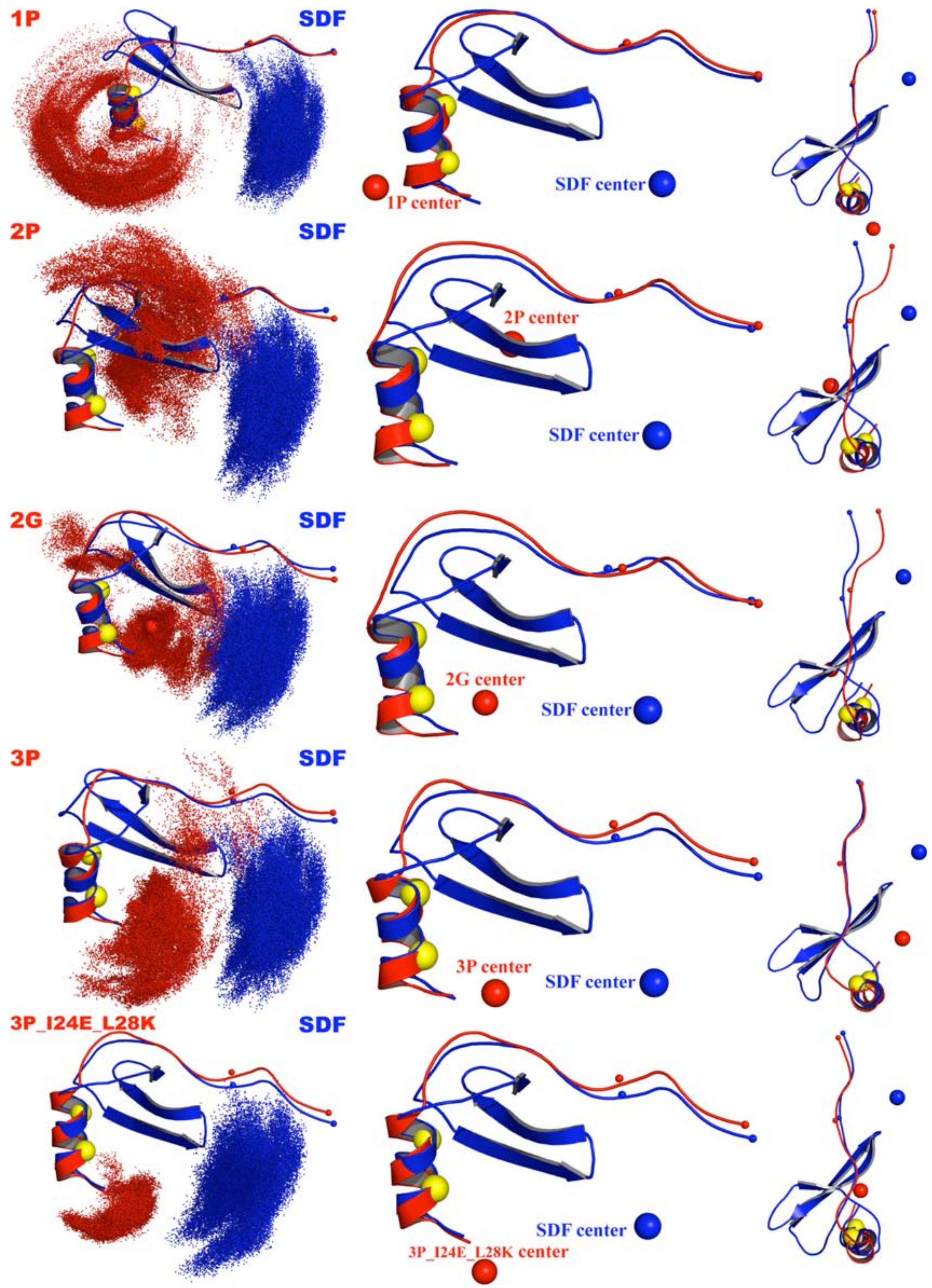


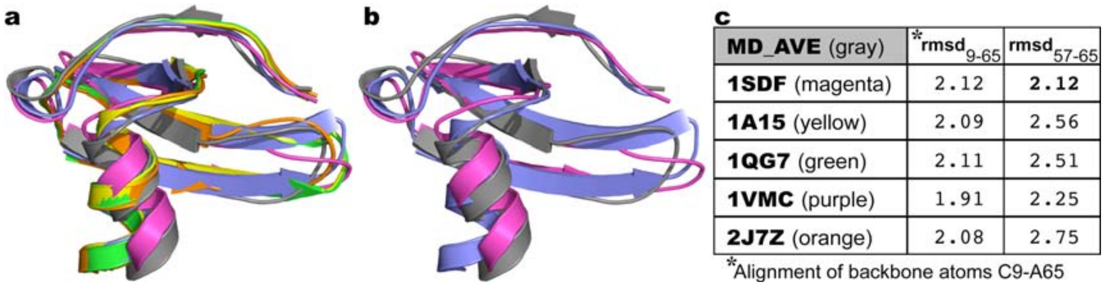
Figure 5.12. Space Sampled by the Center of Mass of the Highly Dynamic N–terminus (K1-R8) Along the MD Simulations. The left panels present comparisons of the space volume explored by the

center of mass of the N-terminus (K1-R8; small spheres, represent the positions of Cα for K1 and R8) of both, SDF–1α (blue) and the simplified analogs (red). The center of the each of the cloud of points is represented as blue (SDF–1α) and red (analogs) spheres, respectively. As reference, the Cα atoms for the equivalent positions I58E and L62K, are rendered as yellow spheres in the analogs structures. These helical positions are facing the β-region in SDF–1α. The overlap for each of the analogs was: 1P (0.000), 2P (0.011), 2G (0.007), 3P (0.001), and 3P_I24E_L28K (0.000). Middle and right panels show two different views of the center of the points explored by the center of mass of the segment K1-R8 in the MD simulations for SDF–1α (blue) and analogs (red).

5.2.3.5 Comparison of Current Crystal and NMR Structures of SDF–1α with the Average Structure from MD simulations

It is worthwhile to compare the average structure of SDF–1α from the MD simulations with the currently available crystal and NMR structures. The initial geometry used in the current study is based on the chain A of the crystal structure, 1A15.pdb [27]. For the production phase (100 ns), the average positions of the backbone atoms were calculated and the structure was compared with the experimentally determined structures from NMR and X-ray crystallographic techniques. The structures used were 1SDF.pdb (NMR) [22], 1A15.pdb (X-ray, 2.20 Å) [27], 1QG7.pdb (X-ray, 2.00 Å) [52], 1VMC.pdb (NMR) [51], and 2J7Z.pdb (X-ray, 1.95 Å) [53]. The results are present in figure 5.13.

The rmsd values indicate that the average structure is closer to the NMR structure 1VMC for the entire segment C9-A65. Interestingly, for the same alignment and as indicated by the $\text{rsmd}_{57\text{-}65}$ values, the position of the α-helix relative to the rest of the structure resembles more the position in the first NMR structure (see Figure 5.13b). Both NMR structures were characterized as monomers.



MD_AVE (gray)	*rmsd$_{9\text{-}65}$	rmsd$_{57\text{-}65}$
1SDF (magenta)	2.12	**2.12**
1A15 (yellow)	2.09	2.56
1QG7 (green)	2.11	2.51
1VMC (purple)	1.91	2.25
2J7Z (orange)	2.08	2.75

*Alignment of backbone atoms C9-A65

Figure 5.13. Comparison of the Average Backbone Structures from the MD Simulations With Current Structures of SDF–1α. (**a**) Superimposed SDF-1α structures: 1SDF.pdb (magenta) [22], 1A15.pdb (yellow) [27], 1QG7.pdb (green) [52], 1VMC.pdb (purple) [51], and 2J7Z.pdb (orange) [53]. (**b**) Comparison of average structure from MD simulations (gray), and the two NMR structures, 1SDF.pdb (magenta) and 1VMC (purple). (**c**) Root mean-square deviation of the experimentally determined structure and the average structure from MD simulations. The structural alignment was done using the backbone atoms from residues C9 to A65. Although the rmsd with 1VMC is the lowest considering the segment C9-A65, the positions of the α-helix relative to the rest of the structure resembles more the 1SDF.pdb structure (rmsd of residues W57-A65).

5.2.4 Discussion

Stromal cell–derived factor–1α (SDF–1α or CXCL12α) is a powerful chemoattractant important in orchestrating migratory responses involved in inflamatory reactions. This small 67-residue protein presents a typical chemokine tertiary structure composed of N–terminus, N–loop, a β-sheet region composed of three β strands, and an α-helix at the C–terminus. There are two difulfide bonds that help to maintain the native conformation of SDF–1α. Despite the fact that the disulfide bonds restrain the conformations of part of the structure, the N–terminal segment composed of residues K1 to R8, has been demostrated to be highly flexible and unstructured [22]. Furthermore, it has been demostrated by mutagenesis studies that the N–terminus is crucial for the binding and for signal transduction [22]. Due to the relevance in different physiological processes, the interactions between SDF–1α and the CXCR4 receptor has been extensively studied. The current model of interaction suggests a two-step/two-site mechanism [22,54]. New evidence suggests that the interactions between the two sites are independent of each other, that is, the first interaction of SDF–1α with the N–terminus of the CXCR4 is independent of the second interaction between the N–terminus of SDF–1α and the transmembrane region of CXCR4 [54]. Additionally, it was been demostrated that quaternary structure of SDF–1α is sensitive to a different factors such as pH, specific anions, heparin, and the presence of CXCR4 [64].

The SDF–1α/CXCR4 system has be investigated as a therapeutic target for the treatment of different disorders [40,56]. Particularly relevant is its possible role in the treatment of cardiovascular diseases [12,46].

A recent study has shown the possibility to use simplified analogs as an alternative neovasculogenic therapy [57]. Here we used molecular dynamics simulations to get insights into the conformations sampled by native SDF–1α and simplified analogs. Different analysis proposed a possible molecular explanation for the beneficial properties of the simplified analogs in mimicking the properties of native SDF–1α.

5.2.4.1 Conformations of SDF–1α and the Simplified Analogs

The findings from MD simulations using two different approaches to compare the conformations sample by SDF–1α and the simplified analogs, suggest that the two-proline analog 2P outperforms the other simplified analogs in mimicking such conformations. This results are in agreement with the experimental finding using a EPC migration and a CXCR4 activation assays [57].

5.2.4.2 Monomer–Dimer Equilibrium and Possible Reasons of the Enhanced Activity of the 2P Analog Relative to Native SDF–1α

To study the role of posttranslational modifications of the N–terminal domain of CXCR4 and its implication in the recognition of SDF–1α, an engineered covalently locked dimer of SDF–1α (denominated $SDF1_2$) was created to simplified the determination of the interactions between SDF–1α and a peptide derived from the N–terminal domain of CXCR4 [65]. The structure of SDF1_2, determined by NMR, was in close agreement with the wild-type dimer determined by X-ray crystallography [52]. Surprisingly, the covalently locked dimer was able to induce intracellular Ca^{2+} mobilization in human acute monocytic leukemia (TPH-1) cells but had no chemotactic activity. Moreover, SDF1_2 inhibits migration of THP-1 cells in response to wild-type SDF–1α and therefore was suggested to act as a potent partial agonist. Also, the role of the shift in oligomeric state was tested using a SDF–1α mutant (H25R) that remains monomeric at higher concentrations. SDF–1α displays a biphasic dose-dependent chemotactic response that decreases at higher concentrations. Interestingly, the SDF–1α mutant (H25R) was shown to maintain chemotactic responses at higher concentrations than native SDF–1α [65]. The final conclusion was that at low concentrations the monomeric SDF–1α is largely dominant and it is responsible for stimulating chemotaxis, whereas at higher concentrations and perhaps promoted by species such as heparin or CXCR4, the dimeric form starts to be dominant and it is resposible for interference with chemotactic signaling [65].

Additionally, Hiesinger et al. performed similar migration studies in endothelial progenitor cells and found similar results, where a biphasic dose-dependent chemotactic response was observed [57]. In this case, the 2P analog (denominated ESA) displayed chemotactic activity at higher concentrations than native SDF–1α. Following similar rationale than in the SDF–1α mutant (H25R), we speculate that the outperformance of the simplified 2P analog is related with its tendency to remain monomeric at high concentrations. As shown in the X-ray structure of the SDF-1α dimer (1QG7.pdb) [52] and the engineered SDF1_2 (2K01.pdb), the β-sheet region and particularly the first β-strand (23VKHLKIL29), is responsible for a large portion of the protein-protein interactions that stabilize the dimer [64]. In the simplified peptide analogs, the β-sheet region was completely removed and with it, some of the interactions that stabilize the dimer formation. To explain the benefits obtained for the

2P simplified analog of SDF–1α, we hypothesize that by removing the β-sheet region, the monomer-dimer equilibrium was shifted to the monomeric form and therefore, the oligomeric state responsible for inducing chemotaxis is present at higher concentrations than in the case of native SDF–1α.

5.2.5 Conclusions

Molecular dynamics simulations have been used to elucidate atomistic details of SDF–1α and simplified analogs. The results show that in aqueous at physiological conditions of temperature and pressure, the two-proline analog explores similar conformations as native SDF–1α under similar simulation conditions. The agreement between experimental and computational findings suggest that MD simulations could be used not only to provide insights into the molecular mechanisms of this type of systems but also to predict protein behaviours in different experimental assays.

5.3 References

1. Lloyd-Jones D, Adams R, Carnethon M, De Simone G, Ferguson TB, Flegal K, Ford E, Furie K, Go A, Greenlund K, et al.: **Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee**. *Circulation* 2009, **119**:e21-181.
2. Araszkiewicz A, Grajek S, Lesiak M, Prech M, Pyda M, Janus M, Cieslinski A: **Effect of impaired myocardial reperfusion on left ventricular remodeling in patients with anterior wall acute myocardial infarction treated with primary coronary intervention**. *Am J Cardiol* 2006, **98**:725-728.
3. Bolognese L, Carrabba N, Parodi G, Santoro GM, Buonamici P, Cerisano G, Antoniucci D: **Impact of microvascular dysfunction on left ventricular remodeling and long-term**

clinical outcome after primary coronary angioplasty for acute myocardial infarction. *Circulation* 2004, **109**:1121-1126.

4. Ejiri M, Fujita M, Sakai O, Miwa K, Asanoi H, Sasayama S: **Development of collateral circulation after acute myocardial infarction: its role in preserving left ventricular function**. *J Cardiol* 1990, **20**:31-37.
5. Balcells E, Powers ER, Lepper W, Belcik T, Wei K, Ragosta M, Samady H, Lindner JR: **Detection of myocardial viability by contrast echocardiography in acute infarction predicts recovery of resting function and contractile reserve**. *J Am Coll Cardiol* 2003, **41**:827-833.
6. Bolognese L, Cerisano G, Buonamici P, Santini A, Santoro GM, Antoniucci D, Fazzini PF: **Influence of infarct-zone viability on left ventricular remodeling after acute myocardial infarction**. *Circulation* 1997, **96**:3353-3359.
7. Garot P, Pascal O, Simon M, Monin JL, Teiger E, Garot J, Gueret P, Dubois-Rande JL: **Impact of microvascular integrity and local viability on left ventricular remodelling after reperfused acute myocardial infarction**. *Heart* 2003, **89**:393-397.
8. Rigo F, Varga Z, Di Pede F, Grassi G, Turiano G, Zuin G, Coli U, Raviele A, Picano E: **Early assessment of coronary flow reserve by transthoracic Doppler echocardiography predicts late remodeling in reperfused anterior myocardial infarction**. *J Am Soc Echocardiogr* 2004, **17**:750-755.
9. Shimoni S, Frangogiannis NG, Aggeli CJ, Shan K, Quinones MA, Espada R, Letsou GV, Lawrie GM, Winters WL, Reardon MJ, et al.: **Microvascular structural correlates of myocardial contrast echocardiography in patients with coronary artery disease and left ventricular dysfunction: implications for the assessment of myocardial hibernation**. *Circulation* 2002, **106**:950-956.
10. Boodhwani M, Sodha NR, Laham RJ, Sellke FW: **The future of therapeutic myocardial angiogenesis**. *Shock* 2006, **26**:332-341.

11. Atluri P, Woo YJ: **Pro-angiogenic cytokines as cardiovascular therapeutics: assessing the potential**. *BioDrugs* 2008, **22**:209-222.

12. Atluri P, Panlilio CM, Liao GP, Suarez EE, McCormick RC, Hiesinger W, Cohen JE, Smith MJ, Patel AB, Feng W, et al.: **Transmyocardial revascularization to enhance myocardial vasculogenesis and hemodynamic function**. *J Thorac Cardiovasc Surg* 2008, **135**:283-291, 291 e281; discussion 291.

13. Jayasankar V, Woo YJ, Bish LT, Pirolli TJ, Chatterjee S, Berry MF, Burdick J, Gardner TJ, Sweeney HL: **Gene transfer of hepatocyte growth factor attenuates postinfarction heart failure**. *Circulation* 2003, **108 Suppl 1**:II230-236.

14. Jayasankar V, Woo YJ, Pirolli TJ, Bish LT, Berry MF, Burdick J, Gardner TJ, Sweeney HL: **Induction of angiogenesis and inhibition of apoptosis by hepatocyte growth factor effectively treats postischemic heart failure**. *J Card Surg* 2005, **20**:93-9101.

15. Kolakowski S, Berry MF, Atluri P, Grand T, Fisher O, Moise MA, Cohen J, Hsu V, Woo YJ: **Placental growth factor provides a novel local angiogenic therapy for ischemic cardiomyopathy**. *J Card Surg* 2006, **21**:559-564.

16. Pillarisetti K, Gupta SK: **Cloning and relative expression analysis of rat stromal cell derived factor-1 (SDF-1)1: SDF-1 alpha mRNA is selectively induced in rat model of myocardial infarction**. *Inflammation* 2001, **25**:293-300.

17. Yamaguchi J, Kusano KF, Masuo O, Kawamoto A, Silver M, Murasawa S, Bosch-Marce M, Masuda H, Losordo DW, Isner JM, et al.: **Stromal cell-derived factor-1 effects on ex vivo expanded endothelial progenitor cell recruitment for ischemic neovascularization**. *Circulation* 2003, **107**:1322-1328.

18. De La Luz Sierra M, Yang F, Narazaki M, Salvucci O, Davis D, Yarchoan R, Zhang HH, Fales H, Tosato G: **Differential processing of stromal-derived factor-1alpha and stromal-derived factor-1beta explains functional diversity**. *Blood* 2004, **103**:2452-2459.

19. Atluri P, Liao GP, Panlilio CM, Hsu VM, Leskowitz MJ, Morine KJ, Cohen JE, Berry MF, Suarez EE, Murphy DA, et al.: **Neovasculogenic therapy to augment perfusion and preserve viability in ischemic cardiomyopathy**. *Ann Thorac Surg* 2006, **81**:1728-1736.

20. Woo YJ, Grand TJ, Berry MF, Atluri P, Moise MA, Hsu VM, Cohen J, Fisher O, Burdick J, Taylor M, et al.: **Stromal cell-derived factor and granulocyte-monocyte colony-stimulating factor form a combined neovasculogenic therapy for ischemic cardiomyopathy**. *J Thorac Cardiovasc Surg* 2005, **130**:321-329.

21. Davis DA, Singer KE, De La Luz Sierra M, Narazaki M, Yang F, Fales HM, Yarchoan R, Tosato G: **Identification of carboxypeptidase N as an enzyme responsible for C-terminal cleavage of stromal cell-derived factor-1alpha in the circulation**. *Blood* 2005, **105**:4561-4568.

22. Crump MP, Gong JH, Loetscher P, Rajarathnam K, Amara A, Arenzana-Seisdedos F, Virelizier JL, Baggiolini M, Sykes BD, Clark-Lewis I: **Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1**. *EMBO J* 1997, **16**:6996-7007.

23. Segers VF, Tokunou T, Higgins LJ, MacGillivray C, Gannon J, Lee RT: **Local delivery of protease-resistant stromal cell derived factor-1 for stem cell recruitment after myocardial infarction**. *Circulation* 2007, **116**:1683-1692.

24. Luo J, Luo Z, Zhou N, Hall JW, Huang Z: **Attachment of C-terminus of SDF-1 enhances the biological activity of its N-terminal peptide**. *Biochem Biophys Res Commun* 1999, **264**:42-47.

25. Fiser A, Do RK, Sali A: **Modeling of loops in protein structures**. *Protein Sci* 2000, **9**:1753-1773.

26. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd, Snoeyink J, Richardson JS, et al.: **MolProbity: all-atom contacts and**

structure validation for proteins and nucleic acids. *Nucleic Acids Res* 2007, **35**:W375-383.

27. Dealwis C, Fernandez EJ, Thompson DA, Simon RJ, Siani MA, Lolis E: **Crystal structure of chemically synthesized [N33A] stromal cell-derived factor 1alpha, a potent ligand for the HIV-1 "fusin" coreceptor**. *Proc Natl Acad Sci U S A* 1998, **95**:6941-6946.
28. Mohle R, Bautz F, Rafii S, Moore MA, Brugger W, Kanz L: **The chemokine receptor CXCR-4 is expressed on CD34+ hematopoietic progenitors and leukemic cells and mediates transendothelial migration induced by stromal cell-derived factor-1**. *Blood* 1998, **91**:4523-4530.
29. Abbott JD, Huang Y, Liu D, Hickey R, Krause DS, Giordano FJ: **Stromal cell-derived factor-1alpha plays a critical role in stem cell recruitment to the heart after myocardial infarction but is not sufficient to induce homing in the absence of injury**. *Circulation* 2004, **110**:3300-3305.
30. Zhao T, Zhang D, Millard RW, Ashraf M, Wang Y: **Stem cell homing and angiomyogenesis in transplanted hearts are enhanced by combined intramyocardial SDF-1alpha delivery and endogenous cytokine signaling**. *Am J Physiol Heart Circ Physiol* 2009, **296**:H976-986.
31. Tang J, Wang J, Yang J, Kong X: **Adenovirus-mediated stromal cell-derived- factor-1alpha gene transfer induces cardiac preservation after infarction via angiogenesis of CD133+ stem cells and anti-apoptosis**. *Interact Cardiovasc Thorac Surg* 2008, **7**:767-770.
32. Tang J, Wang J, Yang J, Kong X, Zheng F, Guo L, Zhang L, Huang Y: **Mesenchymal stem cells over-expressing SDF-1 promote angiogenesis and improve heart function in experimental myocardial infarction in rats**. *Eur J Cardiothorac Surg* 2009, **36**:644-650.

33. Askari AT, Unzek S, Popovic ZB, Goldman CK, Forudi F, Kiedrowski M, Rovner A, Ellis SG, Thomas JD, DiCorleto PE, et al.: **Effect of stromal-cell-derived factor 1 on stem-cell homing and tissue regeneration in ischaemic cardiomyopathy**. *Lancet* 2003, **362**:697-703.

34. Kai H, Ikeda H, Yasukawa H, Kai M, Seki Y, Kuwahara F, Ueno T, Sugi K, Imaizumi T: **Peripheral blood levels of matrix metalloproteases-2 and -9 are elevated in patients with acute coronary syndromes**. *J Am Coll Cardiol* 1998, **32**:368-372.

35. McQuibban GA, Butler GS, Gong JH, Bendall L, Power C, Clark-Lewis I, Overall CM: **Matrix metalloproteinase activity inactivates the CXC chemokine stromal cell-derived factor-1**. *J Biol Chem* 2001, **276**:43503-43508.

36. Peng SB, Peek V, Zhai Y, Paul DC, Lou Q, Xia X, Eessalu T, Kohn W, Tang S: **Akt activation, but not extracellular signal-regulated kinase activation, is required for SDF-1alpha/CXCR4-mediated migration of epitheloid carcinoma cells**. *Mol Cancer Res* 2005, **3**:227-236.

37. Mackay CR: **Chemokines: immunology's high impact factors**. *Nat Immunol* 2001, **2**:95-101.

38. Tilling L, Chowienczyk P, Clapp B: **Progenitors in motion: mechanisms of mobilization of endothelial progenitor cells**. *Br J Clin Pharmacol* 2009, **68**:484-492.

39. Lagerstrom MC, Schioth HB: **Structural diversity of G protein-coupled receptors and significance for drug discovery**. *Nat Rev Drug Discov* 2008, **7**:339-357.

40. Rossi D, Zlotnik A: **The biology of chemokines and their receptors**. *Annu Rev Immunol* 2000, **18**:217-242.

41. Shirozu M, Nakano T, Inazawa J, Tashiro K, Tada H, Shinohara T, Honjo T: **Structure and chromosomal localization of the human stromal cell-derived factor 1 (SDF1) gene**. *Genomics* 1995, **28**:495-500.

42. Gerard C, Rollins BJ: **Chemokines and disease**. *Nat Immunol* 2001, **2**:108-115.

43. Hattori K, Heissig B, Tashiro K, Honjo T, Tateno M, Shieh JH, Hackett NR, Quitoriano MS, Crystal RG, Rafii S, et al.: **Plasma elevation of stromal cell-derived factor-1 induces mobilization of mature and immature hematopoietic progenitor and stem cells**. *Blood* 2001, **97**:3354-3360.

44. Muller A, Homey B, Soto H, Ge N, Catron D, Buchanan ME, McClanahan T, Murphy E, Yuan W, Wagner SN, et al.: **Involvement of chemokine receptors in breast cancer metastasis**. *Nature* 2001, **410**:50-56.

45. Kucia M, Jankowski K, Reca R, Wysoczynski M, Bandura L, Allendorf DJ, Zhang J, Ratajczak J, Ratajczak MZ: **CXCR4-SDF-1 signalling, locomotion, chemotaxis and adhesion**. *J Mol Histol* 2004, **35**:233-245.

46. Takahashi M: **Role of the SDF-1/CXCR4 system in myocardial infarction**. *Circ J* 2010, **74**:418-423.

47. Nagasawa T, Hirota S, Tachibana K, Takakura N, Nishikawa S, Kitamura Y, Yoshida N, Kikutani H, Kishimoto T: **Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1**. *Nature* 1996, **382**:635-638.

48. Tachibana K, Hirota S, Iizasa H, Yoshida H, Kawabata K, Kataoka Y, Kitamura Y, Matsushima K, Yoshida N, Nishikawa S, et al.: **The chemokine receptor CXCR4 is essential for vascularization of the gastrointestinal tract**. *Nature* 1998, **393**:591-594.

49. Zou YR, Kottmann AH, Kuroda M, Taniuchi I, Littman DR: **Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development**. *Nature* 1998, **393**:595-599.

50. Valenzuela-Fernandez A, Palanche T, Amara A, Magerus A, Altmeyer R, Delaunay T, Virelizier JL, Baleux F, Galzi JL, Arenzana-Seisdedos F: **Optimal inhibition of X4 HIV isolates by the CXC chemokine stromal cell-derived factor 1 alpha requires**

interaction with cell surface heparan sulfate proteoglycans. *J Biol Chem* 2001, **276**:26550-26558.

51. Gozansky EK, Louis JM, Caffrey M, Clore GM: **Mapping the binding of the N-terminal extracellular tail of the CXCR4 receptor to stromal cell-derived factor-1alpha**. *J Mol Biol* 2005, **345**:651-658.
52. Ohnishi Y, Senda T, Nandhagopal N, Sugimoto K, Shioda T, Nagal Y, Mitsui Y: **Crystal structure of recombinant native SDF-1alpha with additional mutagenesis studies: an attempt at a more comprehensive interpretation of accumulated structure-activity relationship data**. *J Interferon Cytokine Res* 2000, **20**:691-700.
53. Ryu EK, Kim TG, Kwon TH, Jung ID, Ryu D, Park YM, Kim J, Ahn KH, Ban C: **Crystal structure of recombinant human stromal cell-derived factor-1alpha**. *Proteins* 2007, **67**:1193-1197.
54. Kofuku Y, Yoshiura C, Ueda T, Terasawa H, Hirai T, Tominaga S, Hirose M, Maeda Y, Takahashi H, Terashima Y, et al.: **Structural basis of the interaction between chemokine stromal cell-derived factor-1/CXCL12 and its G-protein-coupled receptor CXCR4**. *J Biol Chem* 2009, **284**:35240-35250.
55. Shoemaker BA, Portman JJ, Wolynes PG: **Speeding molecular recognition by using the folding funnel: The fly-casting mechanism**. *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**:8868-+.
56. Veldkamp CT, Ziarek JJ, Peterson FC, Chen Y, Volkman BF: **Targeting SDF-1/CXCL12 with a ligand that prevents activation of CXCR4 through structure-based drug design**. *J Am Chem Soc* 2010, **132**:7242-7243.
57. Hiesinger W, Perez-Aguilar JM, Atluri P, Marotta NA, Frederick JR, Fitzpatrick JR, 3rd, McCormick RC, Muenzer JR, Yang EC, Levit RD, et al.: **Computational Protein Design to Reengineer Stromal Cell-Derived Factor-1{alpha} Generates an Effective and Translatable Angiogenic Polypeptide Analog**. *Circulation* 2011, **124**:S18-26.

58. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K: **Scalable molecular dynamics with NAMD**. *J Comput Chem* 2005, **26**:1781-1802.

59. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, et al.: **All-atom empirical potential for molecular modeling and dynamics studies of proteins**. *Journal of Physical Chemistry B* 1998, **102**:3586-3616.

60. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML: **Comparison of Simple Potential Functions for Simulating Liquid Water**. *Journal of Chemical Physics* 1983, **79**:926-935.

61. Feller SE, Zhang YH, Pastor RW, Brooks BR: **Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method**. *Journal of Chemical Physics* 1995, **103**:4613-4621.

62. Darden T, York D, Pedersen L: **Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems**. *Journal of Chemical Physics* 1993, **98**:10089-10092.

63. Ryckaert JP, Ciccotti G, Berendsen HJC: **Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes**. *Journal of Computational Physics* 1977, **23**:327-341.

64. Veldkamp CT, Peterson FC, Pelzek AJ, Volkman BF: **The monomer-dimer equilibrium of stromal cell-derived factor-1 (CXCL 12) is altered by pH, phosphate, sulfate, and heparin**. *Protein Science* 2005, **14**:1071-1081.

65. Veldkamp CT, Seibert C, Peterson FC, De la Cruz NB, Haugner JC, Basnet H, Sakmar TP, Volkman BF: **Structural Basis of CXCR4 Sulfotyrosine Recognition by the Chemokine SDF-1/CXCL12**. *Science Signaling* 2008, **1**.

6 Multiple Hindered Rotators in a Gyroscope-Inspired Tribenzylamine Hemicryptophane[1, 2]

6.1 Introduction

Synthetic chemical systems provide means to explore complex phenomena in biological machines [1-8] and also to create novel molecular mechanical components [9]. Designs based on macroscopic devices include brakes, gears, propellers, rachets, turnstiles, rotors, scissors, and most recently gyroscopes [3,10-14]. These molecular systems have been extensively studied in solution, in solid phases, on surfaces and in polymers [15]. One challenge in designing these systems is the required restriction of some molecular degrees of freedom, while allowing specific motions of targeted molecular components [9]. Although synthesis of such molecules can also be challenging, criteria for the construction of molecular gyroscopes have been identified and generally applied: rotary elements (rotators) are attached to a static framework (stator); steric contacts, internal rotation barriers, and interaction with solvent should be minimized to allow low-friction, low-barrier rotary motion; rotating groups are isolated and/or well-separated from each other. To expand upon these criteria, we designed a gyroscope-inspired framework with cyclotriveratrylene (CTV) and trismethylamine as the two-component stator, bridged by three *p*-phenylene rotators (Figure 6.1). Previously, two opposing CTV units have been linked to generate cryptophanes with suitable for host-guest chemistry, as well as biosensing and chiroptical properties [16-31]. CTV has also been employed in supramolecular assemblies, gels and organic microporous polymers [32].

---

[1] This project was a collaboration with the group of Professor Ivan J. Dmochowski at the Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104.

[2] Adapted from Najat S. Khan†, Jose Manuel Perez-Aguilar†, Tara Kaufmann, P. Aru Hill, Olena Taratula, One-Sun Lee, Patrick J. Carroll, Jeffery G. Saven, Ivan J. Dmochowski *J. Org. Chem*. **2011,** 76, 1418-1424. Copyright 2011 American Chemical Society. †N.S.K. and J.M.P-A. contributed equally to this work.

In the current design, a single CTV unit provides a rigid hemicryptophane framework for the synthesis of a novel gyroscope-inspired molecule. Our design incorporates a novel combination of features: (*i*) efficient, high-yielding synthetic scheme, (*ii*) multiple, proximate rotators in one covalently bonded molecular system, (*iii*) exclusion of other molecules and ions from the stator interior that may impede rotator motion, and (*iv*) hindered rotators experiencing friction through exposure to solvent.

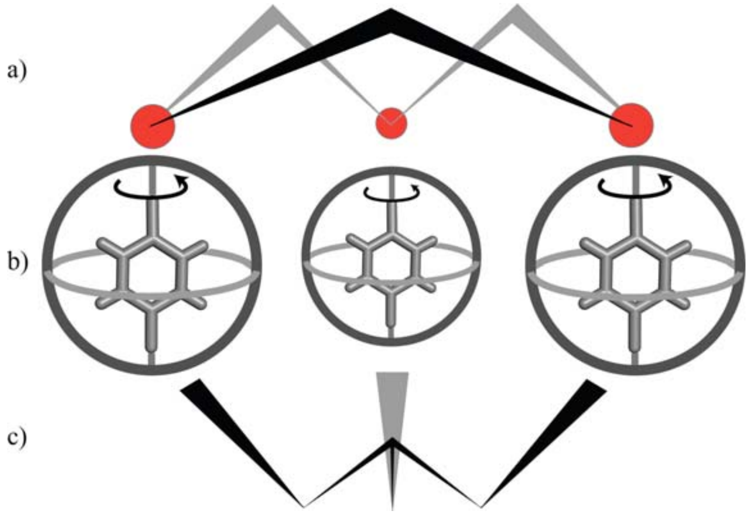


Figure 6.1. Gyroscope-Inspired Tribenzylamine Hemicryptophane (5) Possessing a Rigid Stator (a and c) and Three Rotator Groups (b). Arrows illustrate rotation but are not intended to suggest unidirectionality.

Gyroscope-inspired molecular systems have focused largely on approaching barrierless rotation of isolated or sequestered rotors. One of the first examples (also referred to as a molecular turnstile) was synthesized by Moore and Bedard [33,34]. The creative design included a rigid hexakis(phenylacetylene) framework that preserved the low barrier of rotation about the 1,4-axis of the substituted *p*-phenylene moiety. Garcia-Garibay and co-workers extended these ideas in the construction of amphidynamic crystals [11,15,35-54], where the introduction of bulky substituents creates sufficient space in the lattice framework to allow near barrierless rotation of the central *p*-phenylene group. Furthermore, the *p*-phenylene moiety can be functionalized to create a dipole moment that could be used for controlling motion with an external electric field [11,36-38,42]. Following a different approach, Gladysz and co-workers prepared a series of metal-centered molecular gyroscopes in which the rotator is protected by three-spoke structures as part of the stator. The rotational dynamics of these gyroscopes were studied in solution and their crystal structures indicated sufficient free volume around the rotator to allow low-barrier rotation [9,55-59]. Most recently, Kitagawa *et al*. designed a self-assembled supramolecular gyroscope where the stator is a heterocapsule formed by noncovalent interactions and the rotator is an encapsulated guest [60].

Herein, we report a streamlined synthesis of a gyroscope-inspired tribenzylamine hemicryptophane (**5**, Figure 6.2) involving multiple hindered rotators, where fast rotation is observed by ^{1}H NMR spectroscopy above a critical temperature. Rotations about the 1,4-axis of the three *p*-phenylene rotators encased in a rigid CTV-trismethylamine stator were investigated using ^{1}H variable temperature (VT) NMR spectroscopy and molecular dynamics (MD) simulations.

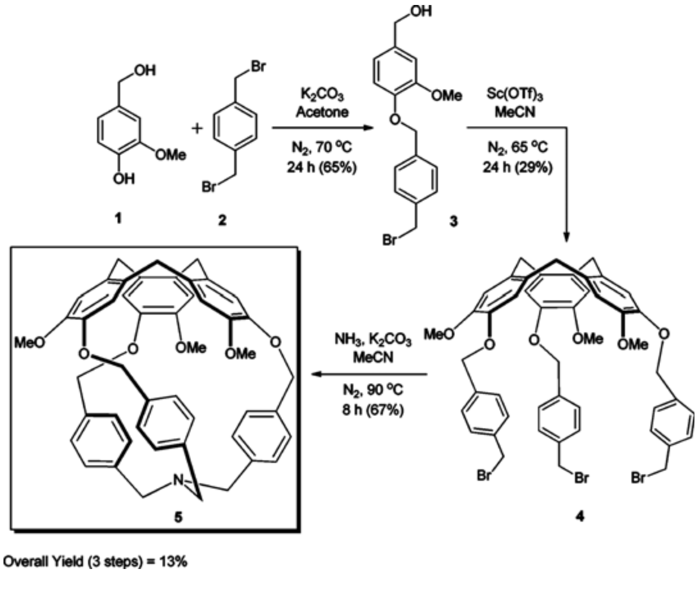
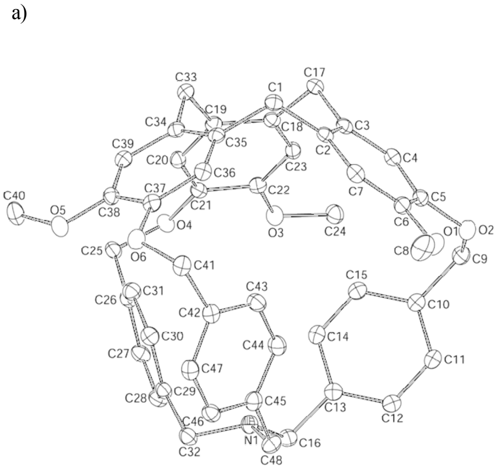


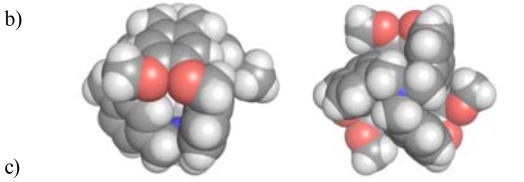
Figure 6.2. Three-Step Synthesis of Gyroscope-Inspired Tribenzylamine Hemicryptophane 5.

6.2 Results and Discussion

6.2.1 Synthesis and X-ray Crystal Structure of Hemicryptophane

In order to synthesize the gyroscope-inspired tribenzylamine hemicrytophane **5,** the linkers (including the rotators) were first cyclized in the rigid CTV framework and then closed with an amine to form the final three-dimensional structure (Figure 6.2). Reaction of commercially available vanillyl alcohol **1** and dibromo-*p*-xylene **2** gave the versatile linker **3** in 65% yield. Cyclization of **3** was achieved with a catalytic amount of Sc(OTf)$_3$ in acetonitrile to afford the "gyroscope scaffold intermediate" **4** in 29% yield. This reaction was based on previous protocols where various 3,4-disubstituted benzyl alcohols were treated with catalytic Sc(OTf)$_3$ to prepare CTV and cryptophane derivatives [18]. Compound **4** was reacted with 7 N NH$_3$ in MeOH to give **5** in 67% yield, with an overall yield of 13% for the three steps. Compound **5** was characterized by solution ^{1}H NMR and ^{13}C NMR spectroscopy, high-resolution mass spectrometry (using the electrospray ionization method), and X-ray crystallography. This short synthetic scheme utilizes mild conditions and results in high overall yields. Moreover, this route provides ample versatility by increasing the number of methylene spacer units or introducing new functional groups on **3** to form rotators with different conformations, rotation barriers, and dipole moments.





c)

N1-C32	1.466 (2) Å	C28-C29-C32-N1	97.7 (2)°
N1-C48	1.464 (2) Å	O2-C9 -C10 -C11	-90.9 (2)°
N1-C16	1.461 (2) Å	O6-C41 -C42 -C47	-89.1 (2)°
C32-C29	1.509 (2) Å	O4-C25 -C26 -C27	-83.1 (2)°
C48-C45	1.512 (3) Å	C14-C13-C16-N1	-61.5 (2)°
C16-C13	1.514 (3) Å	C13-C16-N1-C48	-84.9 (2)°
C32-N1-C48	114.84 (14)°	C30-C29-C32-N1	-77.7 (2)°
C48-N1-C16	113.96 (14)°	C29-C32-N1-C16	-75.4 (2)°
C32-N1-C16	114.45 (14)°	C44-C45-C48-N1	-66.6 (2)°
N1-C32 -C29	108.93 (14)°	C45-C48-N1-C32	-79.4 (2)°
N1-C48 -C45	109.15 (14)°	C12-C13-C16-N1	117.2 (2)°
N1-C16 -C13	110.18 (14)°	C46-C45-C48-N1	107.6 (2)°
O2-C9 -C10 -C15	86.1 (2)°	C13-C16-N1-C32	140.2 (2)°
O6-C41 -C42 -C43	85.4 (2)°	C29-C32-N1-C48	150.05 (13)°
O4-C25 -C26 -C31	93.8 (2)°	C45-C48-N1-C16	145.83 (14)°

Figure 6.3. (a) ORTEP Representations for 5 With Atom Labels. (b) Space-Filling Side and Bottom View of 5. Atom Color Code: C is Gray, O is Red, N is Blue, and H is White. (c) Selected Bond Lengths (Å), Angles (°) and Dihedral Angles (°) of Compound 5.

The X-ray crystal structure of **5** indicates the three rotators adopt a propeller conformation (Figure 6.3). **5** crystallizes in the monoclinic space group $P2_1/c$ and each unit cell consists of four molecules with two of each enantiomer (Figure 6.3). In both enantiomers the *p*-phenylene rotators are oriented edgewise into the interior (angled away from the methoxy group *ortho* to them) with an average dihedral angle of ~ 88.1° for $O\text{-}C_{benzyl}\text{-}C_{phenyl}\text{-}C_{phenyl}$, ~ -68.6° for $N\text{-}C_{benzyl}\text{-}C_{phenyl}\text{-}C_{interior\text{-}phenyl}$ and ~ 107.5° for $N\text{-}C_{benzyl}\text{-}C_{phenyl}\text{-}C_{exterior\text{-}phenyl}$ (see Figure 6.3c). The minimal distance from the phenylic proton on *C44* (pointing into the cage) to the plane described by the next aryl ring (*C10* to *C15*) is relatively small at 3.22 Å. After the van der Waals radii for hydrogen and carbon are included [61], the “clearance” distance is approximately 0.32 Å. The distance from the same proton to the nitrogen atom was even shorter at 2.97 Å, with clearance distance of 0.22 Å when van der Waals radii are included, as the nitrogen is oriented into the interior with an average *C–N–C* bond angle of 114.42° (see Figure 6.3c). This propeller-shaped conformation with a pyramidal nitrogen atom where the lone pair is pointed into the cage is similar to the crystal structure of tribenzylamine, where the rigid CTV is not present [62]. It is believed that this conformation is primarily favored due to steric hindrance [62]. Despite the propeller conformation of the rotators and the positioning of the nitrogen atom into the interior of the cage, rotator motion is expected given the observed clearance distances (Figure 6.3).

Unlike some previously reported porous gyroscopes [58], which encountered barriers to rotation due to the intercalation of solvent molecules (or neighboring molecules in solid state), the small internal volume in **5** should prevent guest encapsulation. Indeed, X-ray crystallography indicated an empty tribenzylamine hemicryptophane, lacking solvent molecules. Small molecules such as helium, dihydrogen, dinitrogen and xenon were not observed to bind to **5** at 1 atm over a range of temperatures, +47 °C to -93 °C. Molecular dynamics simulations were also in agreement with the experimental findings and suggest that these small molecular species should be excluded from the interior. Computational modeling was used to explore effects due to rotation of the *p*-phenylene rotators. GRASP [63] was employed in order to investigate whether a potential interior cavity within **5** emerges with rotation of the rotators. With a probe radius of 1.4 Å, only in improbable, high-energy structures ($\Delta E > 30$ kcal mol^{-1}, where ΔE is the energy relative to the X-ray structure) where the angle of each *p*-phenylene rotator increased by 90° relative to the crystal structure was a cavity identified. This suggests that a cavity of sufficient volume to accommodate small guest molecules is essentially nonexistent, thereby preventing the inclusion of small molecules that may hinder *p*-phenylene rotation.

6.2.2 ^{1}H VT-NMR Experiments with Hemicryptophane 5

In order to investigate the energy barrier of the rotators in solution phase, we performed a ^{1}H VT-NMR study of **5** in CD_2Cl_2 from -93 °C to +27 °C. The NMR spectra in Figure 6.4 indicate that the rotational rate of the *p*-phenylene rotators became slow on the

NMR timescale as the temperature was decreased. At 27 °C, the four protons (labeled H_2, H_3, H_4 and H_5 in Figure 6.4a) on each of the three rotators were split into two doublets at 6.1 ppm (average signal from H_2 and H_4) and 6.9 ppm (average signal from H_3 and H_5). Upon cooling, these doublets became broader and the energy barrier for *p*-phenylene rotation was estimated using equations 6-1 and 6-2 [64], to be 9.2 kcal mol^{-1} from the coalescence temperature (ca. -70 °C; $k_{coalesce} \approx 2300$ Hz). As the temperature was further reduced to -93 °C, a new pair of doublets of equal intensity (three protons each) arose for each doublet that coalesced. This led to splitting of the doublet for H_2 and H_5 and also H_3 and H_4, where H_2 and H_3 shift upfield as they are pointed into the cavity while H_4 and H_5 are oriented away from the cavity. The H_3-H_5 splitting pattern (Figure 6.4b) evidences that *p*-phenylene rotation at -93 °C is slow on the NMR time scale, $t_{rot} > 3$ ms. Similar temperature-dependent behavior was recently reported for a 2,3-dichlorophenylene rotator caged within a polysilaalkane stator [65]. In line with our design, the stator remained rigid throughout the ^1H VT-NMR dataset as indicated by the fact that the integration and splitting pattern of all CTV-trismethylamine proton peaks were constant. It is also interesting to note that protons H_{A1} and H_{A2} remained diastereotopic throughout the ^1H VT-NMR series (Figure 6.4b). All peaks were assigned by ^1H-^1H NOESY experiment (Figure 6.5).

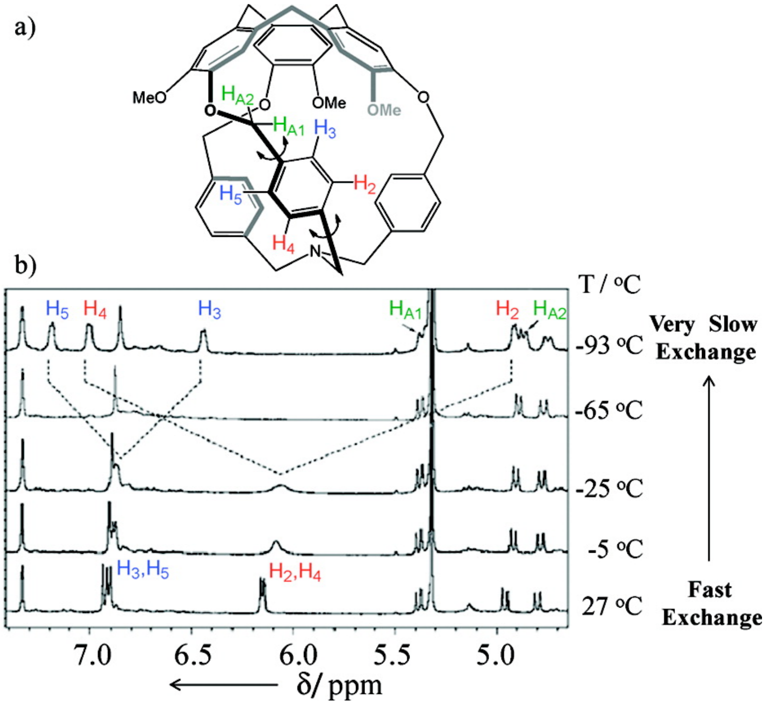


Figure 6.4. (a) Schematic Model of the Molecular Motion of 5. (b) ^{1}H VT-NMR Spectra of 5 Measured in CD_2Cl_2 With 500 MHz Spectrometer.

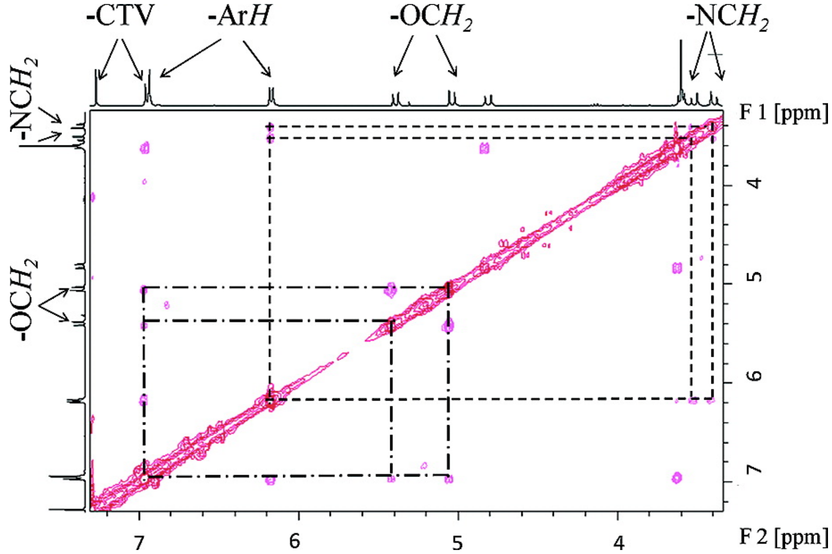


Figure 6.5. ^{1}H-^{1}H NOESY Spectrum of 5 in $CDCl_3$ at 27 °C With 500 MHz Spectrometer to Determine the Assignment of Protons.

6.2.3 Molecular Dynamics Simulations

To investigate the conformational fluctuations of the rotators and stator, MD simulations were carried out on **5** (Figure 6.6). One of the crystal structures was used as the initial structure, and the length of the simulations was 50 ns. Equilibration was confirmed by monitoring relaxation of structural parameters. Three dihedral angles that reflect the conformations of the rotators were selected: ($\alpha \equiv$ *N1-C16-C13-C14*, $\beta \equiv$ *N1-C32-C29-C30* and $\gamma \equiv$ *N1-C48-C45-C44*). The symmetry of the structure yields nearly identical average values of these angles that are in agreement with the values observed in the crystal structure: α = -72.3° ± 10°, β = -71.8° ± 10° and γ = -72.8° ± 10°; the uncertainties (fluctuations) are one standard deviation. The MD simulations at 25 °C indicated limited fluctuations within the structure on the nanosecond time scale (Figure 6.6), and the CTV unit that forms the stator remains highly rigid, in agreement with the NMR results noted above. The *p*-phenylene units are not observed to rotate and instead librate in a manner consistent with a hindered rotor that rotates on a time scale > 100 ns. Therefore, simulations, X-ray, and NMR structural data all agree that the propeller-shaped conformation is highly favored in **5**. High temperatures were artificially employed to observe rotation in the simulations: at 527 °C, rotations of the three *p*-phenylenes are frequent, conformations where the rotators are directed edgewise into the interior (similar to what is seen in the crystal structure) are preferentially populated, and no preferred rotational direction in the three *p*-phenylenes is observed.

To investigate further the *p*-phenylene rotation and interactions among the rotators, the crystal structure of **5** was minimized, and the dihedral angle α was systematically varied (Figure 6.7). The structure was then relaxed via energy minimization, constraining the coordinates of all atoms but those in the two remaining β and γ *p*-phenylene rings. The β-ring is essentially invariant during the rotation. The γ-ring only rotates at most 19.7° to accommodate the 180° rotation of the α-ring. An energy barrier arises due to the steric interaction between the α-ring and γ-ring. Interestingly, the rotational energy barrier estimated from this simplistic and highly constrained calculation (~12 kcal mol^{-1}) is consistent with the experimental value inferred from the ^{1}H VT-NMR experiments (~10 kcal mol^{-1}). The modeling results suggests that, though not entirely independent, the rotations of the α- and γ-rings are only weakly coupled.

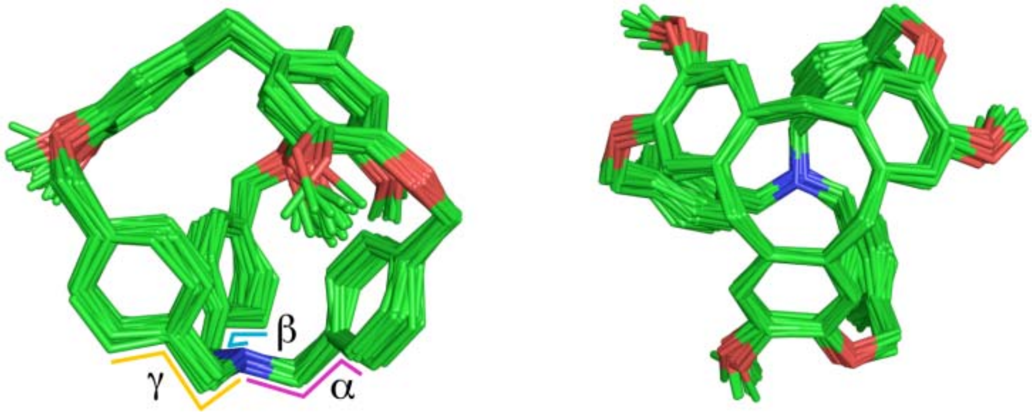


Figure 6.6. Orthogonal Views of 30 Superimposed Structures from the MD Simulation of 5. The dihedral angles are indicated: α (magenta), β (cyan) and γ (yellow).

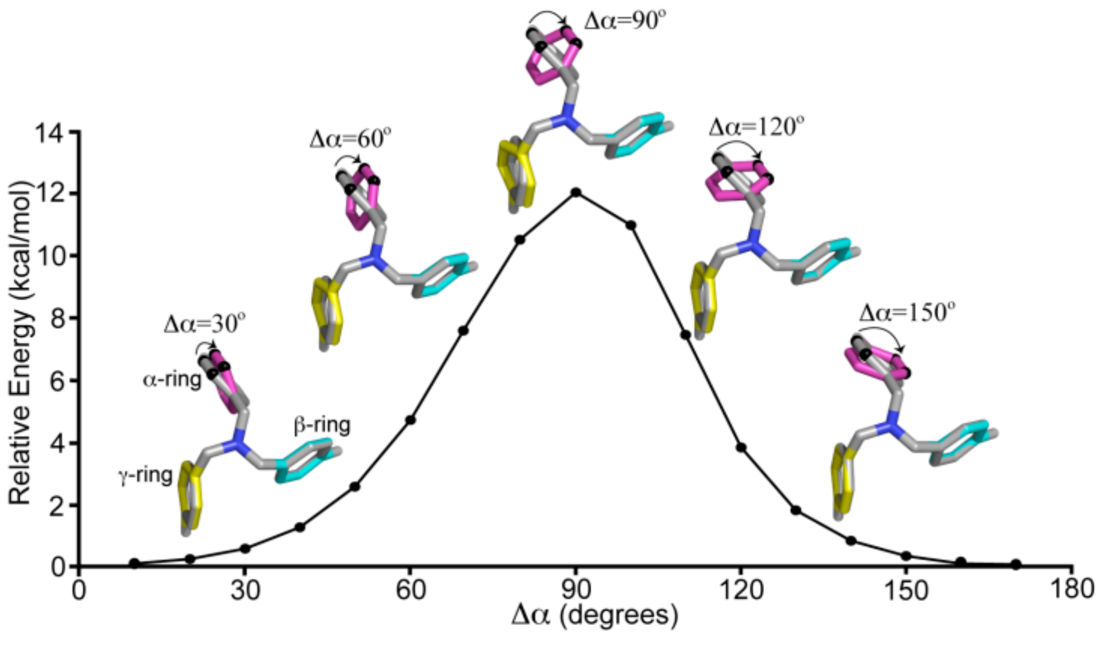


Figure 6.7. Minimized Energy as a Function of Δα, Which is Rotation of the α Dihedral Angle Relative to the Energy Minimized X-ray Structure (Gray). The crystal structure of **5** was minimized, and the angle α was varied in 10° increments, driving rotation of the α-ring (magenta). For each value of Δα, the structure was energy minimized, constraining the coordinates of all atoms but those in the two remaining *p*-phenylene rotators: β-ring (cyan) and γ-ring (yellow). For clarity: the CTV moiety and hydrogens are not shown; two carbons of the α-ring are rendered black.

6.3 Conclusions

In summary, a novel gyroscope-inspired tribenzylamine hemicryptophane **5** was synthesized in three steps in good overall yield using mild conditions. This synthetic route offers the possibility of preparing hemicryptophanes with multiple, proximate rotators, where the molecular properties of the rotators can be varied by changing the length and composition of the linkers used to cyclize the "gyroscope scaffold intermediate" **4**. The compact size of the cavity in this system helps to avoid the inclusion of solvent and gaseous molecules that have the potential to inhibit rotation. ^{1}H VT-NMR data indicate a critical temperature for the onset of rotation on the sub-millisecond time scale and a hindered, dynamic motion of these three rotators at room temperature. The desired rigidity of the CTV stator and the rotator properties of the *p*-phenylenes were corroborated by ^{1}H VT-NMR spectroscopy and MD simulations. As a *p*-phenylene ring rotates, it encounters one of the neighboring *p*-phenylenes, leading to a steric barrier that hinders rotation. Rotation of the rings appears not to be strongly correlated with one another.

The compact size and molecular motions of **5** make it a compelling initial motif from which to engineer unidirectional, potentially coupled rotators for molecular locomotion or transmitting torque. Towards this goal, it may be possible to introduce sterically bulky substituents on the linkers to control the temperatures at which rotation becomes accessible and to favor one direction of rotation. Different substituent groups can also be introduced to create a dipole moment on the rotators, thereby allowing the use of electric fields to explore rapid conformational response in these systems and to control the direction of rotation.

6.4 Experimental Procedures

6.4.1 Calculating from ^{1}H VT-NMR Data the Energy Barrier for *p*-Phenylene Rotation

Equation 6-1 was used to determine the *p*-phenylene rotation rate at the coalescence temperature:

$$k = \frac{\pi \Delta v_0}{\sqrt{2}} \qquad \text{(6-1)}$$

where k is the rate coefficient, $\Delta v_0 = v_A - v_B$ is the chemical shift difference (in Hz) between the two separate signals at slow exchange (in this case at -93 °C or 180 K). The Arrhenius equation (6-2) was used to determine the activation energy, E_a, at the coalescence temperature, T (in K):

$$\ln(k/Hz) = -\frac{E_a}{R}\left(\frac{1}{T}\right) + \ln(A/Hz) \qquad \text{(6-2)}$$

where A is the pre-exponential factor and R is the universal gas constant [64].

6.4.2 Computational Methods

All-atom MD simulations were carried out using NAMD2 [66]. The internal bonded parameters were obtained from AMBER-94 [67], and the nonbonded parameters were proposed in previous studies [68-72]. Simulations were performed in the absence of solvent at 25 °C and temperature was controlled using Langevin dynamics with a damping coefficient of 5 ps^{-1}. The time step of the simulations was 0.5 fs. Relaxation calculations using energy minimization consist of up to 10,000 steps of the conjugate gradient algorithm as implemented in NAMD2; energy was monitored to confirm minimization [66].

GRASP [63] was employed to investigate a potential interior cavity within **5**. Different conformations were generated by manual rotations of the three *p*-phenylenes. The rotations were systematically performed in 10° increments using one of the crystal structures of **5** as reference. A probe radius of 1.4 Å was used for all calculations.

For details about more "Experimental Procedures": Reagents, and Crystal growth and X-ray crystallography, the reader is referred to Najat S. Khan†, Jose Manuel Perez-Aguilar†, Tara Kaufmann, P. Aru Hill, Olena Taratula, One-Sun Lee, Patrick J. Carroll, Jeffery G. Saven, Ivan J. Dmochowski. "Multiple Hindered Rotators in a Gyroscope-Inspired Tribenzylamine Hemicryptophane", *Journal Of Organic Chemistry*, **2011,** 76: 1418-1424. (†Equal contribution) Also the available supporting information.

6.5 References

1. Boyer PD: **Energy, life, and ATP**. *Biosci Rep* 1998, **18**:97-117.
2. Elston T, Wang H, Oster G: **Energy transduction in ATP synthase**. *Nature* 1998, **391**:510-513.
3. Kelly TR, Sestelo JP: **Rotary motion in single-molecule machines**. *Molecular Machines and Motors* 2001, **99**:19-53.
4. Nakanishi-Matsui M, Sekiya M, Nakamoto RK, Futai M: **The mechanism of rotating proton pumping ATPases**. *Biochimica Et Biophysica Acta-Bioenergetics* 2010, **1797**:1343-1352.
5. Rice S, Lin AW, Safer D, Hart CL, Naber N, Carragher BO, Cain SM, Pechatnikova E, Wilson-Kubalek EM, Whittaker M, et al.: **A structural change in the kinesin motor protein that drives motility**. *Nature* 1999, **402**:778-784.
6. Sindelar CV, Downing KH: **An atomic-level mechanism for activation of the kinesin molecular motors**. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**:4111-4116.

7. Terashima H, Kojima S, Homma M: **Flagellar Motility in Bacteria: Structure and Function of Flagellar Motor**. *International Review of Cell and Molecular Biology, Vol 270* 2008, **270**:39-85.

8. Woolley DM: **Flagellar oscillation: a commentary on proposed mechanisms**. *Biological Reviews* 2010, **85**:453-470.

9. Kay ER, Leigh DA, Zerbetto F: **Synthetic molecular motors and mechanical machines**. *Angewandte Chemie-International Edition* 2007, **46**:72-191.

10. Day AI, Blanch RJ, Arnold AP, Lorenzo S, Lewis GR, Dance I: **A cucurbituril-based gyroscane: A new supramolecular form**. *Angewandte Chemie-International Edition* 2001, **41**:275-+.

11. Garcia-Garibay MA: **Crystalline molecular machines: Encoding supramolecular dynamics into molecular structure**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**:10771-10776.

12. Iwamura H, Mislow K: **Stereochemical Consequences of Dynamic Gearing**. *Accounts of Chemical Research* 1988, **21**:175-182.

13. Michl J, Sykes ECH: **Molecular Rotors and Motors: Recent Advances and Future Challenges**. *Acs Nano* 2009, **3**:1042-1048.

14. Skopek K, Hershberger MC, Gladysz JA: **Gyroscopes and the chemical literature: 1852-2002**. *Coordination Chemistry Reviews* 2007, **251**:1723-1733.

15. Karlen SD, Godinez CE, Garcia-Garibay MA: **Improved physical properties and rotational dynamics in a molecular gyroscope with an asymmetric stator structure**. *Organic Letters* 2006, **8**:3417-3420.

16. Bartik K, Luhmer M, Dutasta JP, Collet A, Reisse J: **Xe-129 and H-1 NMR study of the reversible trapping of xenon by cryptophane-A in organic solution**. *Journal of the American Chemical Society* 1998, **120**:784-791.

17. Brotin T, Dutasta JP: **Xe@cryptophane complexes with C-2 symmetry: Synthesis and investigations by Xe-129 NMR of the consequences of the size of the host cavity for xenon encapsulation**. *European Journal of Organic Chemistry* 2003:973-984.

18. Brotin T, Roy V, Dutasta JP: **Improved synthesis of functional CTVs and cryptophanes using Sc(OTf)(3) as catalyst**. *Journal of Organic Chemistry* 2005, **70**:6187-6195.

19. Chambers JM, Hill PA, Aaron JA, Han ZH, Christianson DW, Kuzma NN, Dmochowski IJ: **Cryptophane Xenon-129 Nuclear Magnetic Resonance Biosensors Targeting Human Carbonic Anhydrase**. *Journal of the American Chemical Society* 2009, **131**:563-569.

20. Collet A: *Comp Supramol Chem* 1996, **2**:325-365.

21. Fogarty HA, Berthault P, Brotin T, Huber G, Desvaux H, Dutasta JP: **A cryptophane core optimized for xenon encapsulation**. *Journal of the American Chemical Society* 2007, **129**:10332-+.

22. Gautier A, Mulatier JC, Crassous J, Dutasta JP: **Chiral trialkanolamine-based hemicryptophanes: Synthesis and oxovanadium complex**. *Organic Letters* 2005, **7**:1207-1210.

23. Gosse I, Dutasta JP, Perrin M, Thozet A: **A thiophosphorylated hemicryptophane: structure of the toluene inclusion complex**. *New Journal of Chemistry* 1999, **23**:545-548.

24. Hill PA, Wei Q, Eckenhoff RG, Dmochowski IJ: **Thermodynamics of xenon binding to cryptophane in water and human plasma (vol 129, pg 9262, 2007)**. *Journal of the American Chemical Society* 2007, **129**:11662-11662.

25. Hill PA, Wei Q, Troxler T, Dmochowski IJ: **Substituent Effects on Xenon Binding Affinity and Solution Behavior of Water-Soluble Cryptophanes**. *Journal of the American Chemical Society* 2009, **131**:3069-3077.

26. Martinez A, Robert V, Gornitzka H, Dutasta JP: **Controlling Helical Chirality in Atrane Structures: Solvent-Dependent Chirality Sense in Hemicryptophane-Oxidovanadium(V) Complexes**. *Chemistry-a European Journal* 2010, **16**:520-527.

27. Mynar JL, Lowery TJ, Wemmer DE, Pines A, Frechet JMJ: **Xenon biosensor amplification via dendrimer-cage supramolecular constructs**. *Journal of the American Chemical Society* 2006, **128**:6334-6335.

28. Schlundt A, Kilian W, Beyermann M, Sticht J, Gunther S, Hopner S, Falk K, Roetzschke O, Mitschang L, Freund C: **A Xenon-129 Biosensor for Monitoring MHC-Peptide Interactions**. *Angewandte Chemie-International Edition* 2009, **48**:4142-4145.

29. Seward GK, Wei Q, Dmochowski IJ: **Peptide-Mediated Cellular Uptake of Cryptophane**. *Bioconjugate Chemistry* 2008, **19**:2129-2135.

30. Spence MM, Rubin SM, Dimitrov IE, Ruiz EJ, Wemmer DE, Pines A, Yao SQ, Tian F, Schultz PG: **Functionalized xenon as a biosensor**. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**:10654-10657.

31. Wei Q, Seward GK, Hill PA, Patton B, Dimitrov IE, Kuzma NN, Dmochowski IJ: **Designing (129)Xe NMR biosensors for matrix metalloproteinase detection**. *Journal of the American Chemical Society* 2006, **128**:13274-13283.

32. Hardie MJ: **Recent advances in the chemistry of cyclotriveratrylene**. *Chemical Society Reviews* 2010, **39**:516-527.

33. Kottas GS, Clarke LI, Horinek D, Michl J: **Artificial molecular rotors**. *Chemical Reviews* 2005, **105**:1281-1376.

34. Bedard TC, Moore JS: **Design and Synthesis of a Molecular Turnstile**. *Journal of the American Chemical Society* 1995, **117**:10662-10671.

35. Cizmeciyan D, Yonutas H, Karlen SD, Garcia-Garibay MA: **H-2 NMR and X-ray diffraction studies of methyl rotation in crystals of ortho-methyldibenzocycloalkanones**. *Solid State Nuclear Magnetic Resonance* 2005, **28**:1-8.

36. Dominguez Z, Dang H, Strouse MJ, Garcia-Garibay MA: **Molecular "compasses" and "gyroscopes." - III. Dynamics of a phenylene rotor and clathrated benzene in a slipping-gear crystal lattice**. *Journal of the American Chemical Society* 2002, **124**:7719-7727.

37. Dominguez Z, Dang H, Strouse MJ, Garcia-Garibay MA: **Molecular "compasses" and "gyroscopes". I. Expedient synthesis and solid state dynamics of an open rotor with a bis(triarylmethyl) frame**. *Journal of the American Chemical Society* 2002, **124**:2398-2399.

38. Dominguez Z, Khuong TAV, Dang H, Sanrame CN, Nunez JE, Garcia-Garibay MA: **Molecular compasses and gyroscopes with polar rotors: Synthesis and characterization of crystalline forms**. *Journal of the American Chemical Society* 2003, **125**:8827-8837.

39. Garcia-Garibay MA: **Molecular crystals on the move: From single-crystal-to-single-crystal photoreactions to molecular machinery**. *Angewandte Chemie-International Edition* 2007, **46**:8945-8947.

40. Garcia-Garibay MA: **Molecular machines: Nanoscale gadgets**. *Nature Materials* 2008, **7**:431-432.

41. Garcia-Garibay MA, Dang H: **Photochemical generation, intramolecular reactions, and spectroscopic detection of oxonium ylide and carbene intermediates in a crystalline ortho-(1,3-dioxolan-2-yl)-diaryldiazomethane**. *Organic & Biomolecular Chemistry* 2009, **7**:1106-1114.

42. Godinez CE, Zepeda G, Garcia-Garibay MA: **Molecular compasses and gyroscopes. II. Synthesis and characterization of molecular rotors with axially substituted bis[2-(9-triptycyl)ethynyl]arenes**. *Journal of the American Chemical Society* 2002, **124**:4701-4707.

43. Godinez CE, Zepeda G, Mortko CJ, Dang H, Garcia-Garibay MA: **Molecular crystals with moving parts: Synthesis, characterization, and crystal packing of molecular gyroscopes with methyl-substituted triptycyl frames**. *Journal of Organic Chemistry* 2004, **69**:1652-1662.

44. Gould SL, Tranchemontagne D, Yaghi OM, Garcia-Garibay MA: **Amphidynamic character of crystalline MOF-5: Rotational dynamics of terephthalate phenylenes in a free-volume, sterically unhindered environment**. *Journal of the American Chemical Society* 2008, **130**:3246-+.

45. Jarowski PD, Houk KN, Garcia-Garibay MA: **Importance of correlated motions on the low barrier rotational potentials of crystalline molecular gyroscopes**. *Journal of the American Chemical Society* 2007, **129**:3110-3117.

46. Karlen SD, Garcia-Garibay MA: **Highlighting gyroscopic motion in crystals in C-13 CPMAS spectra by specific isotopic substitution and restricted cross polarization**. *Chemical Communications* 2005:189-191.

47. Karlen SD, Ortiz R, Chapman OL, Garcia-Garibay MA: **Effects of rotational symmetry order on the solid state dynamics of phenylene and diamantane rotators**. *Journal of the American Chemical Society* 2005, **127**:6554-6555.

48. Khuong TAV, Dang H, Jarowski PD, Maverick EF, Garcia-Garibay MA: **Rotational dynamics in a crystalline molecular gyroscope by variable-temperature C-13 NMR, H-2 NMR, X-ray diffraction, and force field calculations**. *Journal of the American Chemical Society* 2007, **129**:839-845.

49. Khuong TAV, Nunez JE, Godinez CE, Garcia-Garibay MA: **Crystalline molecular machines: A quest toward solid-state dynamics and function**. *Accounts of Chemical Research* 2006, **39**:413-422.

50. Kuzmanich G, Natarajan A, Chin KK, Veerman M, Mortko CJ, Garcia-Garibay MA: **Solid-state photodecarbonylation of diphenylcyclopropenone: A quantum chain process**

made possible by ultrafast energy transfer. *Journal of the American Chemical Society* 2008, **130**:1140-+.

51. Nunez JE, Natarajan A, Khan SI, Garcia-Garibay MA: **Synthesis of a triply-bridged molecular gyroscope by a directed meridional cyclization strategy**. *Organic Letters* 2007, **9**:3559-3561.

52. O'brien ZJ, Karlen SD, Khan S, Garcia-Garibay MA: **Solid-State Molecular Rotors with Perdeuterated Stators: Mechanistic Insights from Biphenylene Rotational Dynamics in Ordered and Disordered Crystal Forms**. *Journal of Organic Chemistry* 2010, **75**:2482-2491.

53. Rodriguez-Molina B, Ochoa ME, Farfan N, Santillan R, Garcia-Garibay MA: **Synthesis, Characterization, and Rotational Dynamics of Crystalline Molecular Compasses with N-Heterocyclic Rotators**. *Journal of Organic Chemistry* 2009, **74**:8554-8565.

54. Rodriguez-Molina B, Pozos A, Cruz R, Romero M, Flores B, Farfan N, Santillan R, Garcia-Garibay MA: **Synthesis and solid state characterization of molecular rotors with steroidal stators: ethisterone and norethisterone**. *Organic & Biomolecular Chemistry* 2010, **8**:2993-3000.

55. Nawara AJ, Shima T, Hampel F, Gladysz JA: **Gyroscope-like molecules consisting of PdX2/PtX2 rotators encased in three-spoke stators: Synthesis via alkene metathesis, and facile substitution and demetalation**. *Journal of the American Chemical Society* 2006, **128**:4962-4963.

56. Shima T, Hampel F, Gladysz JA: **Molecular gyroscopes: (Fe(CO)(3)} and {Fe(CO)(2)(NO)}(+) rotators encased in three-spoke stators; facile assembly by alkene metatheses**. *Angewandte Chemie-International Edition* 2004, **43**:5537-5540.

57. Skopek K, Barbasiewicz M, Hampel F, Gladysz JA: **Three-fold intramolecular ring-closing metatheses involving square-planar platinum complexes with cis-**

phosphorus donor ligands: Syntheses, structures, and properties of parachute-like complexes. *Inorganic Chemistry* 2008, **47**:3474-3476.

58. Wang LY, Hampel F, Gladysz JA: **"Giant" gyroscope-like molecules consisting of dipolar Cl-Rh-CO rotators encased in three-spoke stators that define 25-27-membered macrocycles**. *Angewandte Chemie-International Edition* 2006, **45**:4372-4375.
59. Wang L, Shima T, Hampel F, Gladysz JA: **Gyroscope-like molecules consisting of three-spoke stators that enclose "switchable" neutral dipolar rhodium rotators; reversible cycling between faster and slower rotating Rh(CO)I and Rh(CO)(2)I species**. *Chemical Communications* 2006:4075-4077.
60. Kitagawa H, Kobori Y, Yamanaka M, Yoza K, Kobayashi K: **Encapsulated-guest rotation in a self-assembled heterocapsule directed toward a supramolecular gyroscope**. *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**:10444-10448.
61. Bondi A: **Van Der Waals Volumes + Radii**. *Journal of Physical Chemistry* 1964, **68**:441-&.
62. Iwasaki F, Iwasaki H: **The Crystal and Molecular Structure of Tribenzylamine at -70 C**. *Acta Cryst* 1972, **B28**:3370-3376.
63. Nicholls A, Sharp KA, Honig B: **Protein Folding and Association - Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons**. *Proteins-Structure Function and Genetics* 1991, **11**:281-296.
64. Sandstrom J: **Dynamic NMR Spectroscopy; Academic Press: New York**. 1983.
65. Setaka W, Ohmizu S, Kira M: **Molecular Gyroscope Having a Halogen-substituted p-Phenylene Rotator and Silaalkane Chain Stators**. *Chemistry Letters* 2010, **39**:468-469.
66. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K: **Scalable molecular dynamics with NAMD**. *Journal of Computational Chemistry* 2005, **26**:1781-1802.

67. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA: **A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules**. *Journal of the American Chemical Society* 1995, **117**:5179-5197.

68. Kirchhoff PD, Bass MB, Hanks BA, Briggs JM, Collet A, McCammon JA: **Structural fluctuations of a cryptophane host: A molecular dynamics simulation**. *Journal of the American Chemical Society* 1996, **118**:3237-3246.

69. Kirchhoff PD, Dutasta JP, Collet A, McCammon JA: **Structural fluctuations of a cryptophane-tetramethylammonium host-guest system: A molecular dynamics simulation**. *Journal of the American Chemical Society* 1997, **119**:8015-8022.

70. Kirchhoff PD, Dutasta JP, Collet A, McCammon JA: **Dynamic and rotational analysis of cryptophane host-guest systems: Challenges of describing molecular recognition**. *Journal of the American Chemical Society* 1999, **121**:381-390.

71. Potter MJ, Kirchhoff PD, Carlson HA, McCammon JA: **Molecular dynamics of cryptophane and its complexes with tetramethylammonium and neopentane using a continuum solvent model**. *Journal of Computational Chemistry* 1999, **20**:956-970.

72. Rizzo RC, Jorgensen WL: **OPLS all-atom model for amines: Resolution of the amine hydration problem**. *Journal of the American Chemical Society* 1999, **121**:4827-4836.

7 Final Remarks

7.1 Conclusions

In this dissertation, protein systems of varying degrees of complexity were studied using a variety of computational approaches. Also, similar computational methodologies were utilized to study different molecular systems. The computational design approach based on a statistical entropy-based formalism (described in detail in Chapter 2) was used to designed and redesigned different protein systems [1-3]. This approach represented the core methodology of most all the projects described (Chapter 3). All-atom molecular dynamics simulations were used to complement the computational design methods and to extend our understanding of the protein systems under investigation, e.g., SDF-1α (Chapter 5). Two other techniques were used to better understand some of the studied systems: docking calculations and comparative modeling. Docking calculations were used to identify the location and the different poses of ligand molecules into receptors, particularly G protein-coupled receptors (GPCRs) (Chapter 4). Comparative or homology modeling techniques were extensively used to model the structures of GPCRs, e.g., the human μ opiod receptor (Chapter 4).

The study of the human μ opioid receptor, a G protein-coupled receptor, exemplifies the necessity of using many different computational techniques. In generating water-soluble variants for this important receptor, comparative modeling techniques were required to obtain reliable models necessary for computational protein design. As discussed in Chapter 4, the generation of reliable models of the wild type protein by itself is a complex and subtle problem. Something that was particularly useful to guide our modeling was the information collected from mutagenesis studies. Also extremely important for the modeling was the sequence alignment, which is often the essential determining step in comparative modeling protocols. A final consensus was used based on different pairwise sequence alignments programs (see Chapter 4), and the alignment obtained using BLASTp [4] with the Blosum62 [5] substitution matrix was in close agreement with the consensus alignment. In subsequent modeling projects, the combination BLASTp/Blosum62 is likely to yield suitable alignments. We also found that considering the information from the highly conserved fingerprint residues for the rhodopsin-like GPCR family facilitates the generation of reliable structures. Additionally, the use of docking methodologies could provide another way to corroborate specific protein/ligand interactions and, as a consequence, assist in the determination of reliable models. After generating reliable models, the statistical entropy-based formalism will be used to redesign the membrane protein so as to generate water-soluble variants. To test the robustness of the designed structures, all-atom MD simulations will be carried out in aqueous environment under physiological conditions of pressure and temperature.

During our work we identified tools that were valuable in providing insight and analyzing computational and experimental findings. Particular in the case of the designed water-soluble transmembrane $\alpha 1$ domain from the nicotinic acetylcholine receptor (WSA), secondary structure prediction servers were used to predict the locations and lengths of the helices in the water-soluble analog. Different prediction servers were used (PSIPRED [6], SABLE [7], PORTER [8]), but we found that the protein secondary structure prediction server PORTER (http://distill.ucd.ie/porter/), predicted results that were in closest agreement with the NMR-derived structures of WSA. In addition, during the creation of models for a GPCR (μ opioid receptor), the second extracellular loop (EC2) was problematic in the sense that the sequence identity with the current available structures is generaly low [9]. Using the same servers mentioned above, we predicted the secondary structure of EC2 for the currently available GPCR structures. The results from PORTER for assigning secondary structure were more in agreement with the secondary structures observed in β_2 (PDB accession 2RH1) and β_1 (PDB accession 2VT4) adrenergic receptor, bovine rhodopsin (PDB accession 1U19), A_{2A} adenosine receptor (PDB accession 3EML) and the D_3 dopamine receptor (PDB accession 3PBL).

Loop modeling was a technique that was utilized to generate and assay different length loops that connect the N- and C-termini in our work with the SDF-1α analogs (Chapter 5). Although this approach provides no explicit information about structural stability or dynamics, we found it useful to guide the design of these loops and some of the general conclusions from these systems. The loop building and refinement was done using the program Modeller [10,11]. This approach to candidate loop construction offers an easy and low computational cost technique.

Lastly, for the calculation of cavities in proteins, both CASTp [12] and GRASP [13] were found to be useful. While studying smaller molecular systems such as the case of our work with the gyroscope-inspired tribenzylamine hemicryptophane (Chapter 6), GRASP was the preferred choice in calculating cavity volumes. The reason was that GRASP was capable to calculate the volume of the small central cavity of this system. Programs such as CASTp are designed to work with relatively larger cavities such as those found in proteins.

7.2 Ongoing and Future Directions

We close with a few comments regarding ongoing work and potential new research directions.

Regarding the work done in the case of WSA (chapter 3), it will be interesting to explore the transmembrane structures of this important receptor, including those of the other subunits and oligomers of transmembrane domains.

As previoulsly mentioned, although significant effort was put into modeling the structure of the human μ opioid receptor, the final goal was to use the model to design of water-soluble analogs of this important GPCR. Early experimental results are promising in this regard, and it will be of interest to further explore the extent to which the designed protein is consistent with the expected structurally and functionally related properties. In addition, it will be of interest to vary the numbers of residues mutated and use more than one model in generating designed, water-soluble sequences.

In the case of SDF-1α, the information provided from cell-based assays, i.e., Boyden chamber and CXCR4 activation assays, will be compared with the analysis of the molecular dynamics simulations in an effort to provide a molecular interpretation of the activities of designed sequences. These simplified analogs open new possibilities to generate analogs were the MD simulations could be performed to predict the dynamics and binding properties as well as the extent to which such findings can be extrapolated to understand and specify the chemotactic and perhaps even the angiogenic properties of the peptides. Possible routes to improving the peptides are the removal of site(s) known to be the targets of proteases as well as redesign of the current cysteine residues that in SDF-1α form conserved disulfide bonds.

In the case of the hemicryptophane molecules (Chapter 6), the molecular simulations and modeling were in excellent agreement with the observed conformational dynamics and structures of these molecules. Similar computational methodologies could be used to explore and design other molecules containing internal rotors (gyroscopes) and flexible internal degrees of freedom.

7.3 References

1. Saven JG, Wolynes PG: **Statistical mechanics of the combinatorial synthesis and analysis of folding macromolecules**. *Journal of Physical Chemistry B* 1997, **101**:8375-8389.
2. Zou JM, Saven JG: **Statistical theory of combinatorial libraries of folding proteins: Energetic discrimination of a target structure**. *Journal of Molecular Biology* 2000, **296**:281-294.

3. Kono H, Saven JG: **Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure**. *Journal of Molecular Biology* 2001, **306**:607-628.

4. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**:403-410.

5. Henikoff S, Henikoff JG: **Amino acid substitution matrices from protein blocks**. *Proc Natl Acad Sci U S A* 1992, **89**:10915-10919.

6. Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, Jones DT: **Protein annotation and modelling servers at University College London**. *Nucleic Acids Res* **38**:W563-568.

7. Adamczak R, Porollo A, Meller J: **Combining prediction of secondary structure and solvent accessibility in proteins**. *Proteins* 2005, **59**:467-475.

8. Pollastri G, McLysaght A: **Porter: a new, accurate server for protein secondary structure prediction**. *Bioinformatics* 2005, **21**:1719-1720.

9. Worth CL, Kleinau G, Krause G: **Comparative sequence and structural analyses of G-protein-coupled receptor crystal structures and implications for molecular models**. *PLoS One* 2009, **4**:e7011.

10. Sali A, Blundell TL: **Comparative protein modeling by satisfaction of spatial restraints**. *Journal of Molecular Biology* 1993, **234**:779-815.

11. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A: **Comparative protein structure modeling using Modeller**. *Curr Protoc Bioinformatics* 2006, **Chapter 5**:Unit 5 6.

12. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J: **CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues**. *Nucleic Acids Res* 2006, **34**:W116-118.

13. Nicholls A, Sharp KA, Honig B: **Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons**. *Proteins* 1991, **11**:281-296.