

## **Distribution Agreement**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

---

Ashutosh Suhas Jogalekar

---

Date

Part I: Conformational Analysis of  
Bioactive Molecules in Solution

Part II: Structure-Based Design of Selective Inhibitors of  
Cyclin-Dependent Kinase 7

By

Ashutosh Suhas Jogalekar  
Doctor of Philosophy

Chemistry

---

Dennis C. Liotta, Ph.D.  
Advisor

---

James P. Snyder, Ph. D.  
Advisor

---

Justin P. Gallivan, Ph. D.  
Committee Member

---

Fredric M. Menger, Ph. D.  
Committee Member

Accepted:

---

Lisa A. Tedesco, Ph.D.  
Dean of the Graduate School

---

Date

Part I: Conformational Analysis of  
Bioactive Molecules in Solution

Part II: Structure-Based Design of Selective Inhibitors of  
Cyclin-Dependent Kinase 7

By

Ashutosh Suhas Jogalekar  
B. S., M.S., University of Pune

Advisors:  
Dennis C. Liotta, Ph.D.  
James P. Snyder, Ph.D.

An abstract of  
A dissertation submitted to the Faculty of the Graduate School of Emory University  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy  
in Chemistry

2009

# Abstract

## Part I: Conformational Analysis of Bioactive Molecules in Solution

## Part II: Structure-Based Design of Selective Inhibitors of Cyclin-Dependent Kinase 7

By Ashutosh S. Jogalekar

**Part I:** The biological activity of flexible organic molecules crucially depends on their conformation. Knowledge of binding conformation of druglike molecules can help in understanding the fundamental factors operating in protein-ligand interactions and can pave the way toward improvement of biological properties by synthetic modification. Current methods used for determining conformation of flexible, rapidly interconverting molecules in solution provide average data and are therefore inadequate. *NAMFIS* is a method that combines average conformational data from Nuclear Magnetic Resonance (NMR) spectroscopy with extensive force field based conformational searching and derives Boltzmann populations for such flexible molecules in solution. We use this method to investigate the solution conformational profiles of Dictyostatin and Discodermolide, two important potential anticancer molecules binding to the protein tubulin. Using known experimental data, we investigate prior hypotheses regarding their conformations and explore the relation between their solid-state, solution-state and protein bound conformations to construct models of these molecules in the tubulin

binding. The analysis reveals dissimilar and unique conformational behavior in both cases.

Using the same technique, we also investigate conformational profiles for Noscapine, another tubulin binding molecule. A study of conformations in the crystal structures and in solution leads to the observation of an unusual pseudoaxial conformational preference for a methyl group in a six membered ring. Finally, NAMFIS analysis has been applied to previous data for cyclic depsipeptides called Stevastelins that possess immunosuppressant and phosphatase inhibitory activity. The analysis reveals the presence of several previously unobserved unusual conformations that may contribute to the biological activity of the molecule.

The amyloid A $\beta$  (16-22) peptide forms diverse and highly organized structures induced by small changes in physicochemical conditions. The peptide is of interest as a part of the larger amyloid A $\beta$  (1-42) peptide implicated in Alzheimer's disease and forms tubes and fibers under only slightly different conditions of pH. Using molecular dynamics (MD) simulations, molecular mechanics calculations, and structural data obtained from a variety of experiments done by Emory researchers, we investigate the factors responsible for self-assembly. Single-point mutants are utilized to study the contribution of cross-strand pairing to amyloid assembly.

**Part II:** Cyclin-dependent kinase 7 (CDK7) is a protein kinase that plays a crucial role in both phosphorylation of key substrate kinases in the cell cycle as well as transcriptional

activation, and has been implicated in breast cancer. Using a structure-based design strategy, we have extensively collaborated with synthetic chemists and biologists at Imperial College London and have designed BS-181, the first potent and highly selective inhibitor of CDK7 that demonstrates antitumor activity.

Part I: Conformational Analysis of  
Bioactive Molecules in Solution

Part II: Structure-Based Design of Selective Inhibitors of  
Cyclin-Dependent Kinase 7

By

Ashutosh Suhas Jogalekar  
B. S., M.S., University of Pune

Advisors:  
Dennis C. Liotta, Ph.D.  
James P. Snyder, Ph.D.

A dissertation submitted to the Faculty of the Graduate School of Emory University  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy  
in Chemistry

2009

*Dedicated to my parents*



# Acknowledgements

Many individuals provided support, help and encouragement during my time in graduate school and I deeply appreciate their presence.

To Dr. Liotta and Dr. Snyder, two of the best advisors that anyone could find. Not only did they teach me scientific thinking of the highest caliber but more importantly, they taught me through their own example how to be considerate, forthcoming, balanced and patient during interactions with colleagues, teachers and students. The lessons they imparted about always knowing the basics and about questioning assumptions I will carry with me throughout my career. They molded me from being just a student to, hopefully, being a scientist and student.

To Dr. Gallivan, Dr. Menger and Dr. Kindt. Their constructive criticism throughout my time here and their lesson about taking intellectual ownership of my projects are always going to be valuable. I would also like to thank Dr. McDonald for his advice and help during the first year of graduate school.

To Dr. Shaoxiang Wu and Dr. Bing Wang for patiently and carefully teaching me NOESY spectroscopy. To Anita, Donna and Bill from the library for sharing my love of books and helping me get access to almost any written source I requested.

To all my friends and colleagues in the Liotta/Snyder group. Our lab has been one of the friendliest and most helpful labs I have ever encountered. Andy, Ana, Pahk, Matt, Jim, Serdar, Haipeng, Mi-Sun, Jin, Yutao and now Erik and Pieter; thanks a lot for sharing the journey and for always offering much-needed help and support. It was wonderful working with all of you. I would especially like to thank Andy and Ana for sharing the same experiences and for all their help. Many thanks also to Cindy, Onix and Dulce for being the 'housekeeping genes' of our organization.

To Omar; who deserves special mention for sharing the triumphs and woes of graduate school with me right from the first day and first class. He made my life at Emory entertaining, made me truly appreciate Beethoven and left me with memories of hilarious and enlightening moments to be treasured for many years. I have thoroughly enjoyed all of our dinners, Starbucks excursions and conversations; philosophical, scientific, musical and casual.

To my collaborators; Dr. Simak Ali, Prof. Tony Barrett, Prof. Charles Coombes, Prof. Dennis Curran, Damodaran Krishnan, Prof. Dave Lynn, Yan Liang and Anil Mehta who personified close-to-ideal collaborators, always keeping an open mind and sharing every important idea. The CDK7 project would never have gotten its wings off the ground without the friendly folks from across the Atlantic. Thanks to Dave for the ‘kitchen chats’ on the fourth floor. I also want to acknowledge the persistent efforts of Dr. David Lankin Dr. Aiming Sun and Dr. Jaeki Min in acquiring noscapine crystal structures and NOESY spectra. And thanks finally to Basti and Jan who during the course of the project became not just excellent collaborators but great friends.

To all my friends in Atlanta who provided support and encouragement during my stay here. Thanks first and foremost to Jean and Lou, my ‘American parents’ who provided unforgettable love and support. They opened their hearts and their home to me and were always there with me through hard times. Now I can be sure I will always have a family in this country. The same thanks are due to Michael, Edda, Bob, Mona, Steve and Laura, truly wonderful people and friends who are now like family; I will always cherish the time I spent with them. To Chenchu, Chris and James who are my best friends outside school and with whom I spent so many enjoyable weekends during which I learnt a lot. Thanks to all of them for their encouragement and support and for being there during hard times. Special thanks are due to Chenchu for exemplifying a passion for books, reading and knowledge that I have never seen before. Our talks about science, philosophy and history and our trips to bookstores have provided me with much emotional support and inspiration and have sustained me during difficult times.

To Manasi; for all her love, support and help. She has always been there for me through happy as well as sad times in spite of the distance between us, and has always made sure I never give up. The emotional strength she provided has been crucial and her intelligence and practical mind continue to amaze me. She has kept me honest and determined and for this I am grateful.

To my sister, my brother-in-law and their family who have provided so much love, support and encouragement for so long. They have been a constant and reassuring presence in my life, and their practical advice during irrational and uncertain times has been extremely important. I know they will always be there for me. I also appreciate the support of family, friends and teachers from far away Pune, India. Especially acknowledged are my uncles Suresh and Anil and their families, my aunt Sudha and her family, Anand, Yashodhan, the two Neeleshs, Ritwik, Abhay, Sae, Ashlesha, Anagha, Prof. C. K. Desai, Prof. R. K. Pathak and Prof. S. R. Gadre. Also thanks to my relatives in New Jersey- my uncle Satish, my cousin Omkar and their families- for their help and support.

To the writers and scientists who through their writings have provided me with so much intellectual and moral inspiration; they have taught me more than I realize. I would especially like to acknowledge the teachings and words of Hans Bethe, Freeman Dyson, Robert Oppenheimer, Richard Feynman, John L. Casti, Carl Sagan, E. O. Wilson, Constantine P. Cavafy, Arthur Ryder, James Watson, Richard Dawkins, James Gleick and Richard Rhodes. I have often seen the world through your eyes, and I am sure you will continue to motivate.

And finally to my parents, Aai and Baba; the greatest teachers and people in my life who have given me all I could ask for and much, much more that can never be expressed in words. As college professors throughout their career, they inculcated a tremendous atmosphere of learning and inquiry in our home, which imparted lessons about both the sciences and the humanities. I will always be grateful to them for allowing me to pursue

my myriad interests and for developing a love of science and reading in me at a very early age.

My mother was among the kindest, most intelligent and patient people I know. She taught me two crucial lessons for life; to enjoy the small and sundry things that life has to offer, and to always be considerate, patient and on good terms with everyone, to be gentle even in criticism.

My father has one of the finest and most intelligent scientific minds I have encountered and I will be eternally indebted to him for constantly driving home through his own example the importance of curiosity, critical questioning, statistics, balanced reasoning and a sense of wonder about both the physical and the human world. I will carry the lessons he and my mother have taught me all my life, and I can only hope I live up to their expectations and exemplify the dreams they dreamt for me.

To them I dedicate this modest work.

# Table of Contents

## Chapter 1: Molecules, Motion and Life

1.1: Introduction.....	1
1.2: NMR and molecular structure.....	2
1.3: Conformation and drug design.....	4
1.4: NMR and conformation.....	6
1.5: Augmenting NMR data with computational methods.....	10
1.6: NAMFIS.....	12
1.7: Problems with purely theoretical methods for studying conformation .....	18

## Chapter 2: Dictyostatin and Discodermolide

2.1: Microtubules, tubulin and cell division.....	22
2.2: Dictyostatin and Discodermolide.....	28
2.3: Dictyostatin.....	32
2.4: Dictyostatin conformations.....	38
2.5: NAMFIS conformations in DMSO-d <sub>6</sub> and CD <sub>3</sub> OD.....	45
2.6: Dictyostatin conformations in the tubulin binding pocket.....	48
2.7: Outlook and conclusions.....	52
2.8: Discodermolide.....	53
2.9: Conformations of DDM in DMSO-d <sub>6</sub> .....	56

2.10: Simplified DDM constructs.....	63
2.11: The tubulin-bound conformation of DDM.....	67
2.12: Conclusions and outlook.....	71
2.13: Perspective: Conformations of tubulin-binding agents.....	74

## **Chapter 3: Noscapine**

3.1: Introduction.....	77
3.2: NOESY spectra of noscapine.....	78
3.3: Noscapine x-ray crystal structures.....	80
3.4: Conformer classification.....	81
3.5: Noscapine hydrochloride conformations in solution.....	82
3.6: Neutral noscapine conformers in solution.....	87
3.7: Comparison of solution and solid-state conformers.....	93
3.8: Summary and outlook.....	94

## **Chapter 4: The Stevastelins**

4.1: Cyclic peptides.....	96
4.2: The Stevastelins.....	100
4.3: Stevastelin conformations.....	103
4.4: Effect of d-Serine on conformations.....	113
4.5: Energetic analysis.....	115
4.6: Conclusions and outlook.....	118

## **Chapter 5: Selective Inhibitors of Cyclin-Dependent Kinase 7**

5.1: Cancer.....	122
5.2: Phosphorylation.....	126
5.3: Protein kinases.....	127
5.4: Kinase inhibitors in medicine.....	129
5.5: Kinase structure.....	133
5.6: Cyclin-dependent kinases and CDK7.....	136
5.7: Design of selective CDK7 inhibitors.....	141
5.8: Computational modeling of BS-181.....	142
5.9: Biological studies with BS-181.....	152
5.10: Rationalization of CDK7 selectivity and implications.....	157
5.11: Future directions.....	162

## **Chapter 6: Self-assembly of A $\beta$ (16-22)**

6.1: Alzheimer's disease.....	165
6.2: Amyloid A $\beta$ and Alzheimer's disease.....	167
6.3: Self-assembly of A $\beta$ (16-22).....	173
6.4: Cross-strand pairing and $\beta$ -sheet assembly.....	187
6.5: Summary and future directions.....	195

## **Chapter 7: A Bird's Eye View of Computer-Aided**

### **Drug Design**

7.1 Drugs and rational drug discovery.....	200
7.2 Structure-based drug design and CADD.....	203
7.3: Virtual screening.....	205
7.4: Pose prediction.....	208
7.5: Binding affinity prediction.....	209
7.6: Models, computers and drug discovery.....	211



# List of figures

## Chapter 1

**Figure 1.1:** The three dominant conformations of two substituents X and Y around a single bond: from left to right, *gauche-*, *gauche+* and *anti*.....4

**Figure 1.2:** A pair of protons existing in two dominant, rapidly interconverting conformations. The intermediate conformation is an average and is a virtual structure..9

**Figure 1.3:** 2D representation of Taxol (left) and the tubulin-bound conformation of Taxol called T-Taxol (right) which is a minor conformation in solution as found by NAMFIS.....15

## Chapter 2

**Figure 2.1. Top:** Microtubules in human osteosarcoma cancer cells. Microtubules are shown in red, chromatin is in blue and centromeres are in green. **Bottom:**  $\alpha$  and  $\beta$  tubulin dimers assemble to form microtubules that undergo assembly and disassembly at two ends.....23

<b>Figure 2.2:</b> Interaction of diverse tubulin-binding molecules (vinblastine, colchicine and paclitaxel) at diverse binding sites on microtubules.....	24
<b>Figure 2.3:</b> Molecules binding to tubulin.....	25
<b>Figure 2.4:</b> Binding mode of epothilone A in tubulin.....	27
<b>Figure 2.5:</b> Dictyostatin (DCT) and discodermolide (DDM), with the red parts indicating similar substitution patterns.....	28
<b>Figure 2.6. a.</b> Proposed set of conformations (s-cis and s-trans) for the C16-C24 segment of DCT based on NMR data, specifically NOESY correlations <b>b.</b> Proposed overlap of dominant conformation of DCT with x-ray conformation of DDM.....	33
<b>Figure 2.7:</b> Comparison of the Paterson et al. MM2* global minimum <b>2</b> (gold) and <b>3</b> (green), a structure derived from the $^3J_{H-H}$ -derived dihedral angles. For <b>3</b> , a 4.9 Å gap separates two bound carbons in <b>2</b> . Structure <b>3</b> also sustains a 1.6 Å cross-molecule steric clash between a hydroxyl oxygen and a methyl carbon.....	38

**Figure 2.8:** Top three NAMFIS conformers: best fit (NAMFIS-1, 11%), NAMFIS-2 (11%) and NAMFIS-3 (9%). The latter is predicted to enjoy an internal hydrogen bond. The pair below superposes **2** (orange) with NAMFIS-1; all-atom RMSD 4Å.....40

**Figure 2.9:** Overlap of **2** (orange) and the most similar NAMFIS conformation (green, 6<sup>th</sup> best fit, 7%); RMSD 2.7 Å.....41

**Figure 2.10:** C16-C23 fragments for some of the top NAMFIS conformers and the partial NOE data which they satisfy.....43

**Figure 2.11:** Superposition by ring atoms in the C2-C16 region for the most populated ten NAMFIS conformers. RMSDs relative to the top ranked conformer (NAMFIS-1) range from 0.2 to 1.8 Å. The red outline encompasses the C2-C16 segment.....44

**Figure 2.12:** Dictyostatin families in DMSO.....46

**Figure 2.13:** Superposition of similar conformations in DMSO and methanol: From left- DMSO-1 (20%) and methanol-2 (11%) (RMSD 2.3 Å), DMSO-3 (10%) and methanol-1 (11%) (RMSD 1.6 Å) and DMSO-2 (13%) and methanol-3 (9%) (RMSD 1.8 Å).....48

**Figure 2.14:** Overlap of NMR bound conformation of dictyostatin and DMSO-7 (7%)  
The RMSD is 1.6 Å.....49

**Figure 2.14 b:** Best MM-GBSA binding pose of DMSO 7<sup>th</sup> best fit conformation (7%) in tubulin. The inward looking 16-Me group and Phe 270 in tubulin are rendered as space-filling spheres to depict their proximity. The C13-OH group forms a hydrogen bond with a threonine.....51

**Figure 2.15:** Three families of DDM from the Snyder et al. study in DMSO. a. X-ray structure b. Corkscrew (68%), c. Sickle (21%) and d. Extended (10%).....54

**Figure 2.16:** Superposition of the global minima from the four force fields: AMBER\* (blue), MMFFs (cyan), OPLS2005 (gold) and MM3\* (purple). The lactone is at lower left.....56

**Figure 2.17:** Superposition of the four force field global minima and the crystal structure (green) of discodermolide.....57

**Figure 2.18:** The top 3 NAMFIS conformational families of DDM in DMSO-d<sub>6</sub>. From left: Sickle: 17%, Dome: 17% and X-ray: 57% respectively. The top two sustain a hydrogen bond between the carbamate NH<sub>2</sub> and the C-7 hydroxyl. One other conformer belongs to the Extended family (9%).....58

**Figure 2.19:** Overlap of NAMFIS-5 (blue), NAMFIS-9 (pink), and NAMFIS-12 (green). The average heavy-atom RMSD is 0.5 Å.....59

<b>Figure 2.20:</b> Overlap of the X-ray conformation (green) with NAMFIS 12 (pink) and NAMFIS-5 (blue); RMSD = 0.4 Å.....	60
<b>Figure 2.21:</b> Overlap of the DDM X-ray conformation (green) with carbamate excised fragment global minima. Left: C-7, C-11 and C-17 hydroxyls replaced by methyl groups; right: C-7, C-11 and C-17 hydroxyls removed.....	64
<b>Figure 2.22:</b> Overlaps of the global minima for the C6-C24, C7-C17 and C12-C24 fragments (from left to right) of DDM with the X-ray conformation (green).....	65
<b>Figure 2.23:</b> Overlap of DDM tubulin-bound (golden), X-ray (green) and NAMFIS-10 (blue) conformations.....	68
<b>Figure 2.24:</b> The best binding pose of DDM in tubulin as obtained by GLIDE docking followed by MM-GBSA re-docking. The lactone C3-hydroxyl forms a hydrogen bond with a proline carbonyl. Phe 271 rendered in space-filled form does not have measurable contacts with DDM.....	70
<b>Figure 2.25:</b> Constrained DDM analog MMFFs global minimum (cyan) superposed with DDM bound conformation (green) and DDM MMFF global minimum (pink). RMSD is 0.5 Å.....	73

## Chapter 3

**Figure 3.1:** Noscapiene.....77

**Figure 3.2:** Two molecules of noscapine hydrochloride (1-HCl) in the crystal unit cell. The green spheres are chloride ions which form a bridging interaction between the two pseudoequatorial N-H moieties.....81

**Figure 3.3:** The solid state conformations of noscapine. Left: neutral 1 (blue),  $\angle$  (H1-C1-C9-H9) =  $-66^\circ$ ; Right: 1-HCl (green),  $\angle$  (H1-C1-C9-H9) =  $+78^\circ$ .....82

**Figure 3.4:** The third and fifth NAMFIS conformers (1-HCl-3 and 1-HCl-5) differ only in the orientation of the two OMe groups on the lower benzene ring.....84

**Figure 3.5:** Steric clash in the 1-HCl-3 conformer of protonated noscapine, the pseudoequatorial counterpart of the NAMFIS global minimum 1-HCl-1 (41%). The MMFF N-CH<sub>3</sub>H---H<sub>9</sub> interproton distance is 2.3 Å.....86

**Figure 3.6:** Superposition of the x-ray conformation of neutral noscapine (blue carbons) and the most populated equatorial conformer in DMSO (brown carbons).....90

## Chapter 4

**Figure 4.1:** The Stevastelins.....100

**Figure 4.2:** From left: Representative members of the four dominant NAMFIS families for 2R3S; best fit (19%), 2<sup>nd</sup> best fit (16%), 8<sup>th</sup> best fit (5%) and 10<sup>th</sup> best fit (3%), overlap of NAMFIS-1 and previously proposed conformation (olive, RMSD 0.3 Å)...105

**Figure 4.3:** From left: Representative members of the three dominant NAMFIS families for 2S3S; best fit (18%), 2<sup>nd</sup> best fit (16%) and 6<sup>th</sup> best fit (10%), overlap of NAMFIS-1 and previously proposed conformation (olive, RMSD 0.3 Å).....107

**Figure 4.4:** Representatives of the two NAMFIS families for 2R3R. Left: Best-fit (44%) contributing to the dominant family (86%). Right: 4<sup>th</sup> best fit (9%) contributing to the second-most dominant family (15%). Bottom: Overlap of NAMFIS-1 with previously proposed conformer (RMSD 0.5 Å).....110

**Figure 4.5:** From left: Representative members of the four major NAMFIS families for 2S3R. From left; best fit (60%), 2<sup>nd</sup> best fit (18%), 4<sup>th</sup> best fit (8%) and 5<sup>th</sup> best fit (6%), overlap of NAMFIS-1 and previously proposed conformation (olive, RMSD 0.4 Å)...112

**Figure 4.6:** Superposition of the five NAMFIS conformers for 2S3R D-Ser, with the 4<sup>th</sup> best fit (green) sustaining a hydrogen bond between the Val carbonyl and Thr OH....114

**Figure 4.7:** Representative conformations from the three NAMFIS families for 2R3S D-Ser. From left: best fit (52%), 4<sup>th</sup> best fit (5%) and 6<sup>th</sup> best fit (2%).....115

## Chapter 5

**Figure 5.1:** The inhibitor dasatinib bound to Abl kinase illustrating the key features of kinase structure.....133

**Figure 5.2:** Interactions of ATP with residues in the hinge region and binding pocket of a generic kinase.....134

**Figure 5.3:** The cell cycle.....136

**Figure 5.4:** Activation of CDKs involves a conformational changes induced in helix C $\alpha$  by cyclin binding.....138

**Figure 5.5:** Crystal structure of CDK7 with bound ATP (PDB code: 1UA2). Hydrogen bonds with hinge residues Met and Phe and back residues Lys and Ser are shown as dotted red lines. Phe residue in stick form at the back of the adenine ring is gatekeeper. Hydrogen atoms have been omitted for clarity.....142

**Figure 5.6:** AMSOL Solvation energies of six core motifs based on roscovitine. R<sup>1</sup>, R<sup>2</sup> and R<sup>3</sup> are identical to those in roscovitine.....143



**Figure 5.7:** Lowest energy GlideXP pose for pyrazolopyrimidine **1**. The orange space-filling Phe91 is the gatekeeper residue. Hydrogen bonds with Met94 and Asn141 are displayed as dashed lines.....147

**Figure 5.8:** Glide XP pose for BS181 with hydrogen bonds to Met94, Thr170 and Glu20  
.....149

**Figure 5.9:** BS-181 inhibits the growth of MCF-7 tumors in nude mice. Randomized MCF-7 tumour bearing mice were injected intraperitoneally twice daily with 5mg/kg or 10 mg/kg BS-181, giving a total daily dose of 10 mg/kg/day or 20 mg/kg/day, respectively, over a period of 14 days. Mouse weights were determined daily, tumour volumes being measured every 2 days (top). The change in tumour volume was determined for each animal, as tumour volume relative to the tumour volume of each animal at day 1. The line graphs show the mean tumour volumes for the animals in each treatment group. Error bars represent the standard error of the mean. Asterisks depict the statistical significance of the differences between the control group and each of the BS-181 treatment groups (bottom). Shown are the animal weights, as percentage change relative to the animal weights at day 1. Error bars represent the standard error of the mean.....154

**Figure 5.10:** BS-181 inhibits phosphorylation of CDK7 substrates. Whole cell lysates were prepared from MCF-7 cells treated with BS-181 or Roscovitine for four hours, at the concentrations shown. (A) Immunoblotting was carried out using antibodies for RNA polymerase II, or Pol II phosphorylated at Ser2 or Ser5 in the C-terminal domain. The concentration at which apparent inhibition of PolII phosphorylation by 50% would be achieved was determined following densitometry of immunoblots from three experiments. (B-C) Immunoblotting was carried out as in part A, using the antibodies as labeled.....156

**Figure 5.11:** Packing interactions between gatekeeper residue Phe91 and Lys41 in CDK7 (left) and CDK2 (right) .....158

**Figure 5.12: a.** Differences between binding site interactions of typical PPs, purines and BS-181-like PPs. The red substituent corresponds to the substituent interacting with the gatekeeper and the dotted lines indicate the atoms interacting with the hinge region. **b.** The same differences illustrated by interactions in the binding site.....160

**Figure 5.13:** GLIDE (v 3.5) docking scores for BS-181 analogs with different substituents at the C9 position.....161

## Chapter 6

**Figure 6.1:** Histopathological image of stained amyloid plaques (blue) and tau neurofibrillary tangles (purple) in the brain of a person with hereditary AD.....168

**Figure 6.2:** Proteolytic cleavage of the membrane protein APP by  $\alpha$ -secretase produces fragment C83. Proteolysis by  $\beta$  secretase produces fragment C99 which undergoes cleavage by  $\gamma$ -secretase resulting in formation of A $\beta$ .....170

**Figure 6.3:** Structure of A $\beta$  (16-22) peptide fibers and nanotubes in 40% acetonitrile. TEM micrograph of (a) fibers formed at pH 6 (inset shows twisted fiber dimer, scale bar= 50 nm) and (b) tubes at pH 2. Cryo-etch high-resolution SEM images of nanotubes with homogeneous diameters (scale bar= 250 nm) (c) and hollow cross-sections (d) .....174

**Figure 6.4:** Flipping of central strand in a three-strand  $\beta$  sheet necessitated by a shift in register of the strand by one residue.....176

**Figure 6.5:** Schematic of two adjacent strands in a fiber assembly (left) and tube assembly (right). The proximity of the glutamate and lysine residues in the fiber is apparent.....179

**Figure 6.6:** **a.** Edge view of three-strand beta sheet for fiber (left) and tube (right). **b.** Top (left) and bottom (right) views of lipophilic surfaces for three-strand fiber sheet. **c.** Top (left) and bottom (right) views of lipophilic surfaces for three-strand tube sheet. Blue denotes hydrophilic, brown denotes lipophilic and green denotes neutral.....183

**Figure 6.7:** Differences in stacking between fiber peptides (left) and tube peptides (right).....184

**Figure 6.8: Top** a. Twisted fiber assembly with hydrogen bonding between laminates running perpendicular to long axis. B. Unit cell, illustrating characteristic 5 Å hydrogen bonding and 10 Å intersheet distances. c. Tube assembly with hydrogen bonding parallel to long axis. **Bottom:** Differential polarity of two fiber surfaces illustrating lipophilic (brown), charged (blue) and neutral (green) surfaces.....186

**Figure 6.9:** TEM of self-assembled structures: (a) A $\beta$  (16-22) nanotubes assembled under acidic conditions, (b) A $\beta$  (16-22) fibers assembled under neutral conditions, and (c) nanotubes or fibers formed by A $\beta$  (16-22) V18 congeners under acidic conditions; scale = 100 nm. Insets: The side chain structure of the residue substituted at the 18 position.....189

**Figure 6.10:** Calculation of surface complementarity parameter  $S_c$  for two Connolly surfaces  $P_A$  and  $P_B$  involves the dot product of two unit vectors normal to the surfaces at points  $x_A$  and  $x_B$  ( $n_A \cdot n_B$ ) and the proximity of  $x_A$  and  $x_B$  ( $|x_A - x_B|$ ). The greater the dot product and the smaller the distance between the two points, the better the fit and complementarity between the two surfaces.....196

## Chapter 7

**Figure 7.1:** ROC curve for three different VS scenarios. Completely random performance will give the straight white line (AUC 0.5), an ideal performance (no false positives and all true positives) will give the red line (AUC 1.0) and a good VS algorithm will produce the yellow curve ( $0.5 < \text{AUC} < 1.0$ ).....207

# List of Equations and Tables

## Chapter 1

**Equation 1.1:** Typical terms in a force field corresponding to bond stretching and bending, angle bending, torsional and non-bonded Van der Waals and electrostatic terms.....10

**Equation 1.2:** SSD equation for calculating ‘best fit’ solution between calculated (from conformational search) and observed (from NMR) geometric parameters ( $^3J_{\text{H-H}}$  coupling constants interproton distances).  $d$ s refer to distances,  $J$ s refer to coupling constants and  $w$ s refer to weights chosen to optimize the fit and reflect the confidence in the experimental data.....13

## Chapter 2

**Table 2.1:**  $GI_{50}$  values for taxol, DDM and DCT inhibition of wild-type and mutant tubulin strains. The table indicates that while the potency of taxol for the PTX10 and PTX22 cell lines is attenuated, DCT and DDM are almost equally efficacious against them compared to wild-type tubulin.....30

<b>Table 2.2:</b> Dictyostatin $^3J_{\text{H-H}}$ values, the corresponding dihedral angles and the difference in torsion space between the MM2* GM structure (2) and the experimental structure (3).....	37
---	----

<b>Table 2.3:</b> Percentages of DDM populations in DMSO- $d_6$ and $D_2O$ and their classification into x-ray and non-x-ray families.....	62
--	----

### Chapter 3

<b>Table 3.1:</b> Geometries and Boltzmann energies of the NAMFIS conformers of noscapine hydrochloride salt 1-HCl in $CDCl_3$ .....	83
--	----

<b>Table 3.2:</b> Relative energies of 1-HCl NAMFIS conformers, kcal/mol.....	85
---	----

<b>Table 3.3:</b> Geometries and Boltzmann energies of the NAMFIS conformers of noscapine (1) in DMSO- $d_6$ and $CDCl_3$ .....	88
---	----

<b>Table 3.4:</b> Relative energies of noscapine (1) NAMFIS conformers, kcal/mol.....	91
---	----

### Chapter 4

<b>Table 4.1:</b> Relative force field and ab initio energies for the six NAMFIS global minimum diastereomers.....	117
--	-----

## Chapter 5

**Table 5.1:** List of approved anticancer kinase inhibitors currently on the market.....130

**Table 5.2:** Docking scores and solvation energies of core templates.....144

**Table 5.3:** Comparison of Glide XP docking scores for BS181 in CDK7, 2, 5 and 6.....150

**Table 5.4:** BS-181 is more selective for CDK7 compared to Roscovitine.....151

**Table 5.5:** GI<sub>50</sub> values for BS181 compared to roscovitine in various cell lines.....152

## Chapter 6

**Table 6.1:** Summary of the OPLS-05 and AMBER 94 energy difference between in-register and one-residue-shifted registries of Ab(16-22) and its V18 congeners. For each peptide, the simulation is done with 3-strand  $\beta$ -sheet in antiparallel in-register and antiparallel one-residue-shifted-register. The energy difference is calculated as  $\Delta\Delta E = \Delta E_{\text{in-register}} - \Delta E_{\text{one-residue-shifted}}$  for each peptide, in both force fields.....192

**Table 6.2:** Buried surface area of A $\beta$  (16-22) and its V18 congeners.....194



# List of Abbreviations

A $\beta$ : Amyloid-beta

AD: Alzheimer's disease

CDK: Cyclin-Dependent Kinase

DCT: Dictyostatin

DDM: Discodermolide

MD: Molecular Dynamics

PP: Pyrazolopyrimidines

PTX: Paclitaxel

SAR: Structure-Activity Relationships

# **Chapter 1**

## **Molecules, Motion and Life**

“ There are more things in heaven and earth, Horatio,  
Than are dreamt of in your philosophy”

-Hamlet, William Shakespeare

### ***1.1 Introduction:***

**T**he basis of all life is molecular motion. As the great physicist Richard Feynman said<sup>1,2</sup>:

“If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generations of creatures, what statement would contain the most information in the fewest words? I believe it is the atomic hypothesis (or the atomic fact, whatever you wish to call it) that all things are made of atoms – *little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another.* In that one sentence, you will see, there is an enormous amount of information about the world, if just a little imagination and thinking are applied.” (Emphasis added)

The complex organic and inorganic structures that serve as fuels, building blocks, signaling agents and genetic messengers in biological systems are dynamic assemblies that constantly interact with each other and with the fluid environment in which they are immersed. Their biological effects are often crucially dependent on the motion they undergo. If we want to understand the behavior of these molecules and extract useful information from that behavior that can be converted into practical knowledge such as medical therapies, we need to understand both their structure and dynamics.

Fortunately, a hundred years of development in physics, chemistry and instrumentation have provided us with exquisite tools to study both structure and dynamics of molecules.

In the early days of chemistry, structure determination depended on arduous chemical detective work in which a molecule of unknown structure would be broken down into fragments using chemical methods. It would then take an impressive amount of detective work to mentally reassemble these fragments and deduce their sum total<sup>3</sup>. In the twentieth century, a remarkable set of tools developed initially by physicists converged on chemistry and biology to provide a wealth of structural information. Out of all these methods, two stand out for their lasting importance. The first one, x-ray crystallography, developed by Max von Laue<sup>4</sup>, the Braggs<sup>5</sup> and others provides detailed atomic-level resolution of molecules in the solid state. This technique was first used for small organic molecules, after which a series of post-World War II breakthroughs enabled scientists to apply it to proteins. This development was extremely important and that importance continues to this day<sup>6</sup>. However, molecules in the solid-state are static and we already mentioned their dynamic nature. While x-ray crystal structures are usually a fair representation of molecular structures in living systems, a technique that determined their structure in solution would be of inestimable importance. Fortunately, such a technique was developed in the 1950s- NMR spectroscopy<sup>7</sup>.

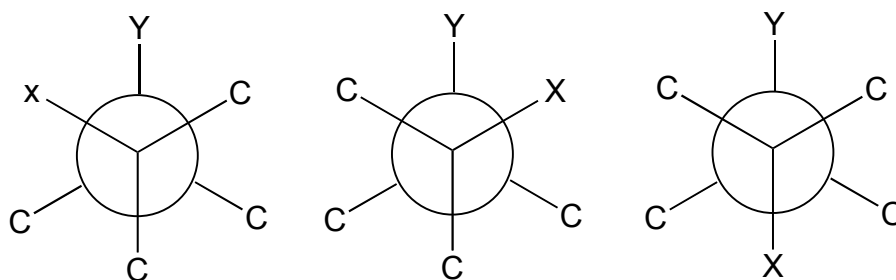
### ***1.2: NMR and molecular structure***

Of all the techniques available for structure determination NMR- a technique that distinguishes atoms in different environments based on their magnetic properties- has distinguished itself as one of the most important. Five Nobel prizes have been awarded to researchers whose work was instrumental in the development and application of this technique. In its early days, NMR spectroscopy made huge contributions to the routine

structure determination of organic molecules. Today no organic synthesis or natural products study proceeds without NMR spectroscopy. Stereochemistry is powerfully determined by NMR for example<sup>7-9</sup>. In the last two decades NMR has also been used to determine the structure of proteins<sup>10</sup>. It has also manifested itself in the shape of Magnetic Resonance Imaging (MRI), a now-routine technique that has made possible enormous strides in medical diagnostics<sup>11</sup>. Complemented with x-ray diffraction, NMR provides a set of tools that has allowed unprecedented insight into the behavior of molecules.

What distinguishes NMR spectroscopy from many other techniques is that while it is very useful for determining *configuration* of chiral centers in molecules, it is also very useful for determining *conformation*. A molecule can exist in several conformations in solution. The relative proportions of these conformations, usually referred to as a Boltzmann population distribution, are determined by their relative free energies. Determination of configuration has always been important for organic chemists, but determination of conformation has special importance for insights into the behavior of flexible molecules, especially those that are used as drugs. Let us see why that is so.

Most drugs are small organic molecules of molecular weight less than 500 daltons that can have several rotatable bonds<sup>12</sup>. About a single rotatable bond there are three dominant conformations; two gauche and one anti as shown in **Figure 1.1** below<sup>8</sup>.



**Figure 1.1:** The three dominant conformations of two substituents X and Y around a single bond: from left to right, *gauche-*, *gauche+* and *anti*

Thus, since one bond has three possible conformations, a molecule with, for instance, ten rotatable bonds will have a maximum of  $3^{10}$  or about 60,000 conformations, a huge number. Many of these conformations may be forbidden by high-energy steric clashes between atoms, but still there will be a considerable number in solution.

### ***1.3: Conformation and drug design***

The concept of rational drug design involves the precise manipulation of biologically active molecules based on a fundamental understanding of their structure and properties. Most drugs are flexible molecules that bind to proteins and modulate their activity<sup>13</sup>. Many of the best selling drugs on the market today such as the ubiquitous aspirin and the cholesterol-lowering atorvastatin for example bind to proteins or enzymes and regulate their activity. When a flexible molecule binds to a protein, it binds in a single conformation. However, the protein has to pay an energy penalty for converting the conformations of the molecule in solution to the bound conformation or otherwise has to extract the conformer resembling the bound one from the solution<sup>13,14</sup>. This penalty is

largely entropic since the molecule loses its degrees of freedom when it binds to the protein<sup>15</sup>. Studies have estimated the average strain energy required to be expended by a ligand to bind to a protein at about 2 kcal/mol<sup>14,16</sup>.

Determining this protein-bound conformation is very important for drug design and is a much-pursued goal in pharmaceutical science. For example, knowing what the bound conformation is, organic chemists can make modifications to the structure to 'lock' the molecule into that conformation and thus avoid paying the entropic penalty, forcing the molecule to be already constrained into the favorable conformation before it binds<sup>13</sup>. If one could have an x-ray crystal structure of every drug bound to every important protein, the problem would be largely resolved. But unfortunately protein crystallization is still an art and it has been very difficult to determine x-ray structures for some of the most important proteins bound to their corresponding ligands. X-ray conformations of protein ligand complexes can also obscure protein motion. Now it has been noted several times that the conformation that actually binds to the protein- termed the 'bioactive' conformation- is usually identical to or close to one among the several in solution<sup>10-12</sup>. Thus, it is clearly valuable to have a list of the conformations of a druglike flexible molecule in solution. A precise correlation between the structure of the bound conformer and its population in solution is not necessary. For example, the bound conformation can be present to the extent of only 2-5% in solution<sup>17</sup>; as long it's a part of the conformational equilibrium, the protein can extract it out of solution.

### ***1.4: NMR and conformation***

For determining conformation of such molecules, NMR spectroscopy provides two very important variables that relate to conformation. One is the proton-proton coupling constant,  $^3J_{\text{H-H}}$ . The coupling constant relates to conformation through the valuable Karplus equation<sup>18,19</sup>; insert the coupling constants into the equation and one derives the dihedral angles about bonds and therefore the conformation. Over the years the equation has been parameterized for several kinds of substituents and bonds<sup>19,20</sup>. Another important variable is the distances between protons. One can acquire them through a powerful technique called NOESY (Nuclear Overhauser Effect Spectroscopy) that maps protons separated by distances of about 5 Å or less<sup>7</sup>. Together, the dihedral angles and the interproton distances define the conformations of the molecule. Taken together, both these variables provide a powerful set of parameters for determining conformation in solution.

Nonetheless, the problem of determining conformation in solution by NMR is a recalcitrant problem. To see why this is so, it is important to understand that the relative populations of equilibrating conformations in solution are governed by their relative free energies in solution. There is a relation between free energy and equilibrium constant that is all-important in this situation;  $\Delta G = -RT \ln K$ , where  $K$  is the equilibrium constant,  $\Delta G$  is the free energy difference,  $R$  is the gas constant and  $T$  is the temperature. A look at the equation tells us that since  $K$  is exponentially dependent on  $\Delta G$ , a small difference in  $\Delta G$  makes a huge difference in  $K$ <sup>8</sup>. For example consider the example of a molecule that can exist in only two conformations in solution. If the free energy difference between them is



only 1.8 kcal/mol (compare this to the strength of a typical carbon-carbon bond; 83 kcal/mol), the lower energy conformer will be about 97% while the higher energy one will be just 3%. Increase the energy difference to 3 kcal/mol and the higher-energy one will be virtually non-existent, with the other one about 99.96%. One can clearly imagine an ensemble of 10-15 conformations, if anything an underestimate in case of highly flexible molecules, in which even the “dominant” conformation would be about 20%, and several other conformations could range from 2-5%, with the energy differences between all these being tiny. NMR cannot detect conformations with a percentage of less than 2-3% because of their vanishingly small concentration in solution; the conformations simply would not provide enough signal strength for NMR to discern their existence<sup>8</sup>.

Thus, determination of conformation by NMR is firstly plagued by thermodynamics; in this case the technique is plagued by its low sensitivity

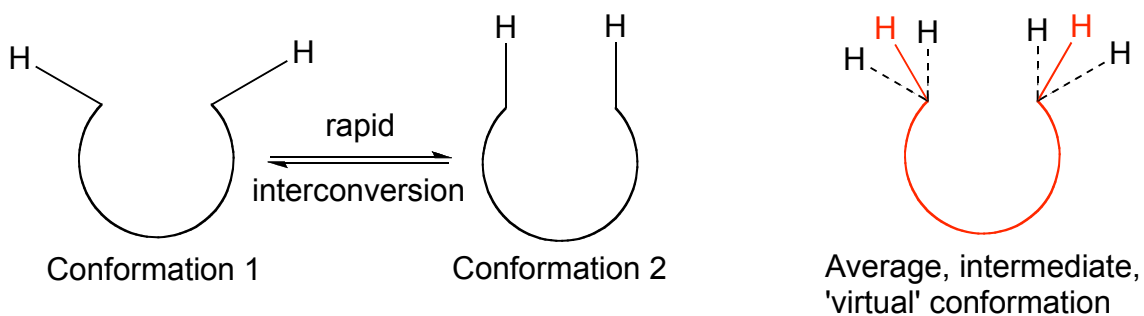
Kinetics poses an even greater constraint for determining conformations using NMR. Rotational barriers between conformations ( $\Delta G^{\ddagger}$ ) can be very small compared to available thermal energy at room temperature. For example, the classic rotational barrier for interconversion in ethane is only 3 kcal/mol<sup>21</sup>. Energy available at room temperature is about 20 kcal/mol, which makes the ethane conformations interconvert very fast. So even for energy barriers that are several kcal/mol, conformational interconversion is usually more than adequate to observe averaging of conformations and consequently all associated parameters- most importantly chemical shifts and coupling constants- in NMR. The resolution time of NMR is on the order of tens of milliseconds, while

conformational interconversion is on the order of tens of microseconds or less. In theory one can go to lower temperatures and 'freeze out' such motions. The resolving power of NMR also depends on the field strength of the magnet. In many such experiments, line broadening at lower temperatures is observed, followed by separation of peaks at the relevant temperature. But consider that even for a barrier as high as 8-10 kcal/mol, NMR usually gives distinct, separate signals for the different conformers only at -100 degrees Celsius. For barriers like those in ethane, the situation would be hopelessly challenging. From an experimental standpoint, this fact implies that sharp, well-defined resonances at room temperature do not indicate lack of conformational interconversion but can simply mean that conformational interconversion is fast compared to the NMR time scale.

Thus, determination of conformation by NMR is also plagued by kinetics; in this case the technique is plagued by low resolution time.

To see in some detail why determining average structures by NMR is flawed, it is constructive to compare NMR with some other resolution techniques. Trying to gauge conformations using NMR is like trying to make out individual blades of a fan through the human eye when it is moving very fast, or trying to photograph a horse race with a camera with a low shutter speed. The former analogy is especially instructive; on observation of a rapidly rotating fan, one sees a disk instead of individual blades because the rotation speed of the blades is greater than the resolution time of the eye. Thus, the disk one sees is an average of all the motions. Similarly as mentioned above, when we use NMR spectroscopy to determine coupling constants and interproton distances, we get

average values for these. In case of the fan, the contiguous disk we see is clearly a ‘virtual’ average structure. Similarly, any structure that one might assign to the average coupling constants and interproton distances from NMR for a flexible molecule will be an average and therefore non-existent, often called a ‘virtual’ structure. As a simple example, consider two protons that exist in two conformations as shown below in **Figure 1.2**. In one they are apart and in another one they are close. They rapidly interconvert between the two orientations. Clearly the average is the intermediate structure which does not exist in reality.



**Figure 1.2:** A pair of protons existing in two dominant, rapidly interconverting conformations. The intermediate conformation is an average and is a virtual structure.

This is an important point. NMR is extremely valuable for getting information on rapidly interconverting conformations in solution. But this information is an average for all those conformations, and assigning a single structure to this average data is inherently flawed. In this sense it is a misnomer to refer to the singular ‘conformation’ for a flexible molecule in solution; one should always talk about ‘conformations’.

Needless to say then, when conformational studies with NMR are attempted, it should always be kept in mind that thermodynamics leads to low populations and kinetics leads to averaging of populations. Clearly, fitting the average NMR data to a single conformation for a flexible molecule is inherently flawed and unrealistic, even when the molecule is relatively constrained. Yet the literature is replete with instances of studies in which a single structure or a single family of structures is fit to the NMR as the sole representative structure in solution<sup>22,23</sup>. On the other hand, there does not seem to be an easy solution to overcoming the problem of low-concentration, rapidly interconverting conformations in solution.

### ***1.5: Augmenting NMR data with computational methods***

Fortunately there is a possible solution. Computational advances in the last three decades have provided several theoretical methods for investigating conformations of organic molecules. One of the most important and widely used methods is molecular mechanics, which uses collections of equations and experimentally derived parameters called force fields to investigate the potential energy surface of a molecule<sup>24,25</sup>. A typical force field will have equations calculating the total energy of a molecule from individual energies for bond stretching, angle bending, dihedral rotations, electrostatic interactions and Van der Waals interactions (**Equation 1.1**).

$$\begin{aligned} E_{total} &= E_{stretch} + E_{bend} + E_{torsion} + E_{non-bonded} \\ E_{non-bonded} &= E_{VdW} + E_{electrostatic} \end{aligned}$$

**Equation 1.1:** Typical terms in a force field corresponding to bond stretching and bending, angle bending, torsional and non-bonded Van der Waals and electrostatic terms.

The equations are based on classical mechanics and treat the atoms and bonds as balls and springs. Because this picture is approximate, experimental parameters are used to augment the equations and make them accurately reproduce molecular geometries and energies. Over the years several different force fields such as MM2, MM3, AMBER, MMFF and OPLS have been developed, and each one of these involves careful parameterization of the equations using data from experiment and high-level quantum mechanics calculations to reproduce the properties of complex organic molecules. Force field parameters are usually designed to be ‘transferable’ so that they can be applied to a wide variety of molecules, and it is not uncommon for modern force fields to have thousands of parameters for simulating a diversity of molecular structures.

Using these force fields, one can do a conformational search of a flexible molecule. This consists of exploring the conformational potential energy surface and finding all possible low-energy conformations of a molecule. In practice it is quite challenging to find all possible conformations, but most often using a good energy window to capture many conformations and using a large number of search steps, one can do reasonably well in mapping the conformational surface<sup>26</sup>. After a conformational search one ends up with a large number of conformations, usually numbering in the thousands.

### ***1.6: NAMFIS***

The key point is that one can now combine this set of conformations with the average NMR data above using deconvolution methods that can dissect that data into individual conformations. The goal is to essentially compare the exhaustively enumerated conformations with the average NMR data and generate a dataset that aims to replicate the Boltzmann population of the conformers in solution. Several such deconvolution methods have been proposed before<sup>27-29</sup>, but a particularly tractable method named NAMFIS has emerged as a facile one in the last decade or so.

NAMFIS stands for NMR Analysis of Molecular Flexibility In Solution. It was developed by researchers at the University of Rome in Italy<sup>30</sup> and transformed into a workable method by researchers at Emory University in Atlanta in the United States. NAMFIS takes as its input two sets of variables; one is the set of average coupling constants and average distances from NMR, and the other is the set of conformations computed through the conformational search. NAMFIS can calculate the distances and coupling constants for all the theoretical structures. To calculate the coupling constants, NAMFIS uses the highly parameterized Haasnoot-Altona version of the Karplus equation<sup>19</sup>. This parameterized six-term equation takes into account substituent electronegativities and orientations of substituents with respect to particular protons.

What NAMFIS then does is to vary the mole fractions of the theoretical conformations, calculate the resulting coupling constants and distances, and compare these weighted parameters to the corresponding experimental data. In effect, by varying the mole

fractions, NAMFIS simulates different conformational populations in solution and then compares the resulting calculated parameters to the average experimental parameters. The fit of the two is expressed as Sum of Squares Differences (SSD), a standard measure of fits between sets of data (**Equation 1.2**).

$$SSD = 1/2 \left( \sum_{i=1}^m [(d_{calc}(i) - d_{exp}(i) / d_{err}(i)]^2 \right) + \left( \sum_{i=1}^m [w(j) \times ({}^3J_{calc}(j) - {}^3J_{exp}(j) / 1.5]^2 \right) + 1/2$$

**Equation 1.2:** SSD equation for calculating ‘best fit’ solution between calculated (from conformational search) and observed (from NMR) geometric parameters ( ${}^3J_{H-H}$  coupling constants interproton distances).  $ds$  refer to distances,  $Js$  refer to coupling constants and  $ws$  refer to weights chosen to optimize the fit and reflect the confidence in the experimental data.

The  $ds$  and  $Js$  refer to distances and coupling constants and the  $ws$  refer to weights that are chosen to reflect the degree of confidence in the data, with a higher weight (empirically chosen to be 3) assigned to the usually more accurately determined coupling constants. Clearly the ‘best-fit’ solution would be the one for which this number is the smallest. The accompanying mole fractions would then represent the ideal ensemble of conformations in solution. Note that there will be several combinations of conformations that could fit the data more or less equally well. NAMFIS chooses the best fit among these. Contrary to methods like simulated annealing<sup>31</sup>, constrained conformational searches or restrained molecular dynamics that constrain a single structure or structural family to fit the geometric NMR data, NAMFIS varies the mole fraction of every

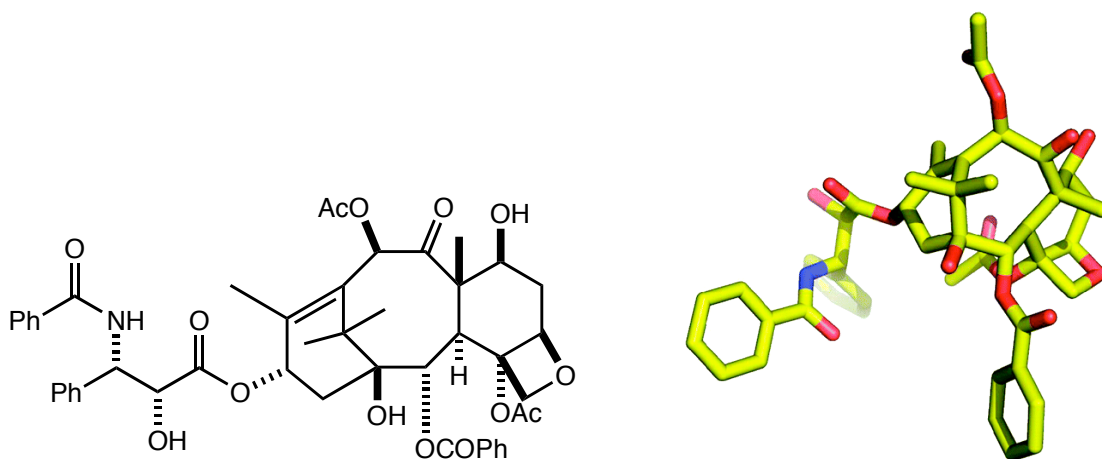
structure from a comprehensive conformational search and fits the resulting ensemble to the NMR data until a best fit for the entire dataset is obtained, thus deriving a Boltzmann conformer population in solution. By performing a purely geometric fit of the two datasets and taking into account the energies implicit in the NMR data, NAMFIS avoids the ambiguities in locating low-energy conformations of flexible, polar molecules resulting from incomplete sampling of conformational space and inadequate representation of solvation<sup>32</sup>. Thus NAMFIS serves as a useful alternative to purely theoretical methods for investigating molecular conformations. Ultimately, NAMFIS derives a whole family of conformations that together satisfy the average NMR data, a much realistic proposition than assigning all the data to a single conformation.

NAMFIS is a powerful tool because it can tease out individual conformations from average NMR data; a goal that we have seen cannot be accomplished by NMR alone. Another significant advantage of NAMFIS is that it derives a relatively small number of structures; 10-20 in all the cases studied until now, compared to the theoretical thousands. It has been used to provide Boltzmann populations for many important molecules and shed light on their bioactive conformations. A few representative examples attesting to the utility of NAMFIS follow.

NAMFIS analysis was used to probe the conformations of Taxol. This molecule is a key therapeutic against cancer and kills cancer cells by binding to the important protein tubulin and disrupting cell division. The structure of the bound conformation of Taxol in the tubulin binding pocket was elucidated using electron crystallography, a technique



similar to x-ray diffraction that uses electrons. NAMFIS analysis generated a dozen Taxol conformers in two solvents. The bioactive conformation of Taxol turned out to be one that was present to the extent of only 2% in solution (**Figure 1.3**)<sup>17</sup>. This fact underscores the value of NAMFIS in finding minor conformations in solution that would not be found by NMR. Similar NAMFIS analysis was also applied to another anticancer molecule binding to tubulin called epothilone<sup>33</sup>. The observation of the bioactive conformer as a minor solution conformer was also underscored in this study.



**Figure 1.3:** 2D representation of Taxol (left) and the tubulin-bound conformation of Taxol called T-Taxol (right) which is a minor conformation in solution as found by NAMFIS (Figure from Kingston<sup>34</sup>).

Peptides pose particular challenges for conformational analysis. They often present polar functionalities in highly flexible motifs and peptide structures can sometimes be misassigned by combining the entirety of the average data in a single common motif such as an alpha helix, most commonly with techniques like simulated annealing or restrained

molecular dynamics. In another instructive NAMFIS application, the technique was used to determine the family of solution conformations for a 5-residue peptide that was claimed to form an alpha helix in solution<sup>35</sup>. Confidence in the existence of the helix came from the observation of characteristic and contiguous signals in the NMR data and the fact that the two ends of the peptide were tethered together by a metal to apparently constrain the peptide in a helical conformation. NAMFIS analysis indicated however that the NMR data could be much better satisfied by a collection of conformations that did not include the alpha helix even as a minor conformation. The principal conclusion from this study was that average NMR data may give the illusion of a dominant conformation, and secondly that even constraining a molecule may not actually force it in a given conformation. Note that NAMFIS does not have to necessarily conclude the absence of such a conformation; more typically in case of bioactive molecules, we have located the bioactive conformation as a minor solution conformation (*vide infra*).

Finally, NAMFIS analysis was used to explore the solution conformations of geldanamycin and radicicol, two inhibitors of the heat shock protein (Hsp90) that is intimately involved in diseases like cancer. In this case the study was particularly illuminating as crystal and protein-bound structures were also available for the two molecules so that one could compare the results in solution. NAMFIS analysis for NMR data obtained in CDCl<sub>3</sub> indicated bioactive conformations for geldanamycin and radicicol to be present in solution with percentages of 4% and 21% respectively<sup>36</sup>.

As seen from the case studies above, not only does NAMFIS allow us to represent the NMR data more meaningfully with a set of multiple conformations, but it also can be insightful in correcting or modifying previous conclusions. Also note that in many cases, the bioactive conformation deduced through other experimental techniques was very similar to a minor conformation in solution found by NAMFIS, and not the dominant conformation in solution. As seen above, it would not be possible to find this low-population conformation using NMR alone. Another virtue of NAMFIS is its speed. Most calculations themselves take not more than several minutes, while conformational searching for a complex molecule typically may take a few hours; the principal bottleneck is in the acquisition of the NMR data and especially in the recording of accurate NOESY spectra that enables the extraction of interproton distances.

In the most general case, one can question whether the protein-bound bioactive conformation is truly present in solution. While there is no reason to necessarily assume this, most of the NAMFIS studies have indicated this to be so<sup>36</sup>. In addition there is a theoretical basis for trusting this correspondence. Studies have shown the average strain energy of ligands to be around 2-3 kcal/mol<sup>37</sup>, which means that the global minimum solution conformation of a ligand may be at most about 3 kcal/mol higher in energy than the bound conformation, a fact that will ensure its presence in solution. But even if this energetic hierarchy does not strictly hold, the conformations from NAMFIS nonetheless provide a more tractable set of structures for exploring the active site of the protein as explained in the next section.

### ***1.7: Problems with purely theoretical methods for generating viable bioactive conformations***

There are many cases in which the bioactive conformation of a compound is not known because of experimental constraints such as problems in crystallizing the protein-ligand complex. To try to circumvent this problem, a common computational technique used in drug design today is molecular docking<sup>38,39</sup>. In this technique, a program places or ‘docks’ thousands of conformations for a bioactive ligand generated by a conformational search in the active site. It then performs a conformational and translation search of the ligand in the binding site and determines the interaction of each conformation with the atoms in the protein; this includes interactions such as hydrogen bonding, electrostatic interactions, and hydrophobic interactions. The program then calculates the sum total for the favourable and unfavourable interactions for each conformation and assigns a number called the ‘score’ which denotes how well the conformation binds. It then ranks the scores from best to worst, and the best scoring conformation is posited as the bioactive conformation of the ligand.

It would be extremely valuable if such fast theoretical methods could provide bioactive conformations without doing a structure determination experiment. In practice, both the conformations from a conformational search and the scores determined by docking them are notoriously fickle<sup>39</sup>. Conformational searches for example can generate spurious high-energy structures and docking programs may sometimes score these high-energy structures favourably, leading to false positives<sup>40</sup>. One can also have low-energy structures that score unfavourably, leading to false negatives. In general the absence of a

sound reference standard presents a severe constraint in assigning theoretically generated structures as ‘low’ or ‘high’ in energy since this assignment depends on the particular method used and its accompanying parameters, with the result that different methods give divergent results. But NAMFIS significantly bypasses this problem by deriving a relatively small representative set of experimental low-energy structures that can serve as docking candidates. Since the structures satisfy the NMR data, they have to be low-energy by definition otherwise they would not exist in solution. Also, the complexity of the problem is reduced by a factor of hundred or so since NAMFIS delivers a dozen or so structures compared to the thousands generated from a conformational search.

There will undoubtedly be future applications of NAMFIS in determining conformations of medicinally important molecules in solution and generating hypotheses for bioactive conformations. As hinted above, one drawback of NAMFIS pertains not to the method itself but with the data acquisition for it. Doing a conformational search is a standard protocol automated in several molecular modeling programs; the choice of number of search steps, method of searching and energy window must nonetheless be judicious. However, the other part of the NAMFIS input, namely acquiring accurate NOESY data from NMR studies and extracting distances from it, is a non-trivial exercise within the domain of skillful NMR operators<sup>41</sup>. Nonetheless, the efforts spent in this data acquisition are well worth the wealth of conformational data that one can unearth for the purposes of understanding the behavior of molecules in both solution and inside protein active sites. NAMFIS is a reminder of the power of both NMR spectroscopy and modern computational chemistry to shed light on molecular structures and their behavior in living

systems. There is adequate assurance that such techniques will increasingly continue to aid us in fundamental understanding of molecules and their involvement in health and disease.

# **Chapter 2**

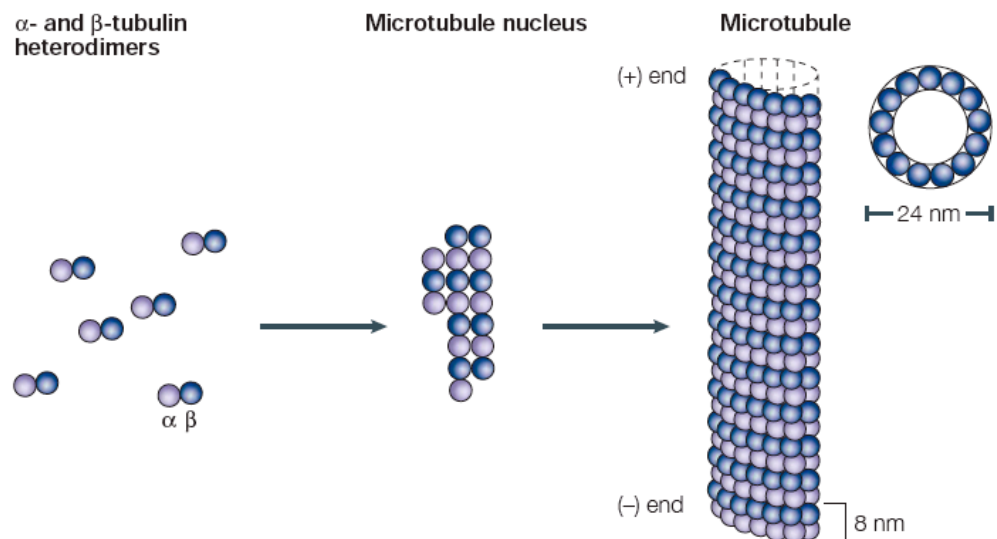
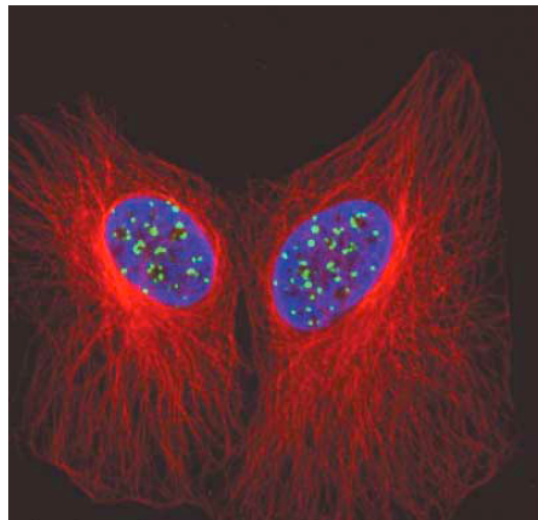
## **Dictyostatin and Discodermolide**

“ God is subtle, but malicious he is not”

-Albert Einstein

### 2.1: Microtubules, tubulin and cell division

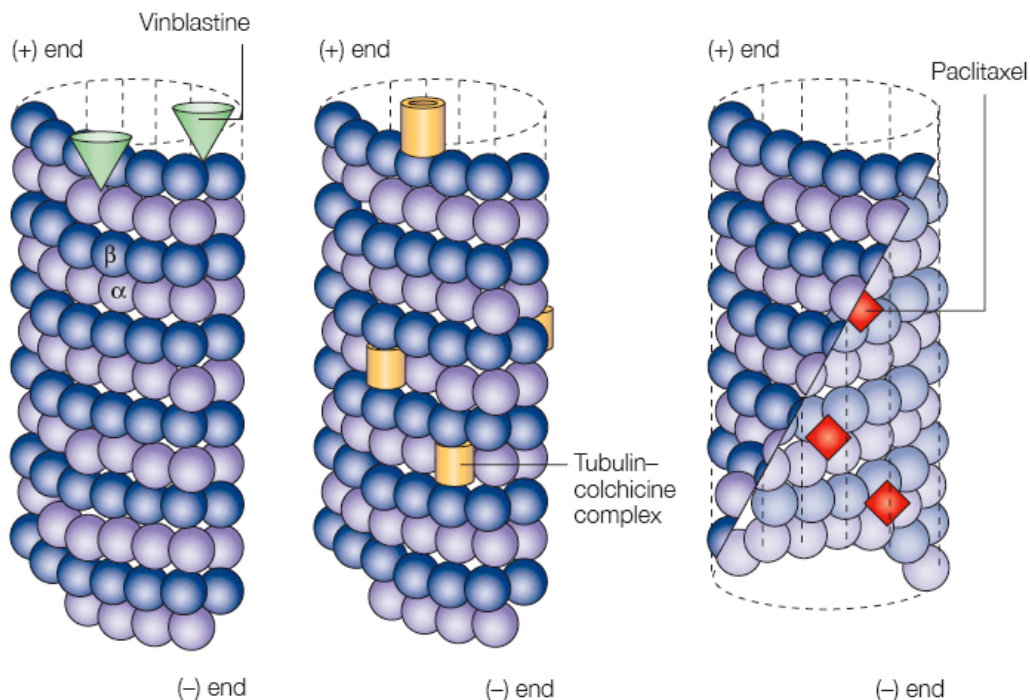
**M**icrotubules are dynamic protein structures whose assembly and disassembly determines key events during cell division. They exist as fleeting polymeric assemblies that direct spindle formation in metaphase<sup>42</sup>. Microtubules are composed of  $\alpha$  and  $\beta$  dimers of the protein tubulin and undergo constant transient assembly at one end (+) and disassembly at the other (-), a process termed ‘dynamic instability’<sup>43</sup> (**Figure 2.1**).





**Figure 2.1. Top:** Microtubules in human osteosarcoma cancer cells. Microtubules are shown in red, chromatin is in blue and centromeres are in green. **Bottom:**  $\alpha$  and  $\beta$  tubulin dimers assemble to form microtubules that undergo assembly and disassembly at two ends (Figures from Jordan et al.<sup>43</sup>)

Drugs that target microtubules are of particular interest in the fight against cancer<sup>43</sup>. Paclitaxel (PTX, Taxol®), a molecule that was isolated from the Pacific Yew tree, has become one of the best selling drugs in the world for treating breast cancer and colon cancer. PTX and related drugs bind to the beta-subunit of the alpha-beta dimer of the protein tubulin that constitutes the building block of microtubules, and modulates microtubule dynamics by either stabilizing or destabilizing their assembly. Some compounds like PTX, the epothilones<sup>44</sup>, discodermolide<sup>45</sup> and dictyostatin<sup>46</sup> stabilize microtubule formation, facilitate their polymerization and prevent their transient disassembly while others like colchicine<sup>47</sup> and vinblastine<sup>48</sup> inhibit their formation and polymerization. These molecules are thought to exercise their complex actions by binding to different sites on microtubules (**Figure 2.2**).

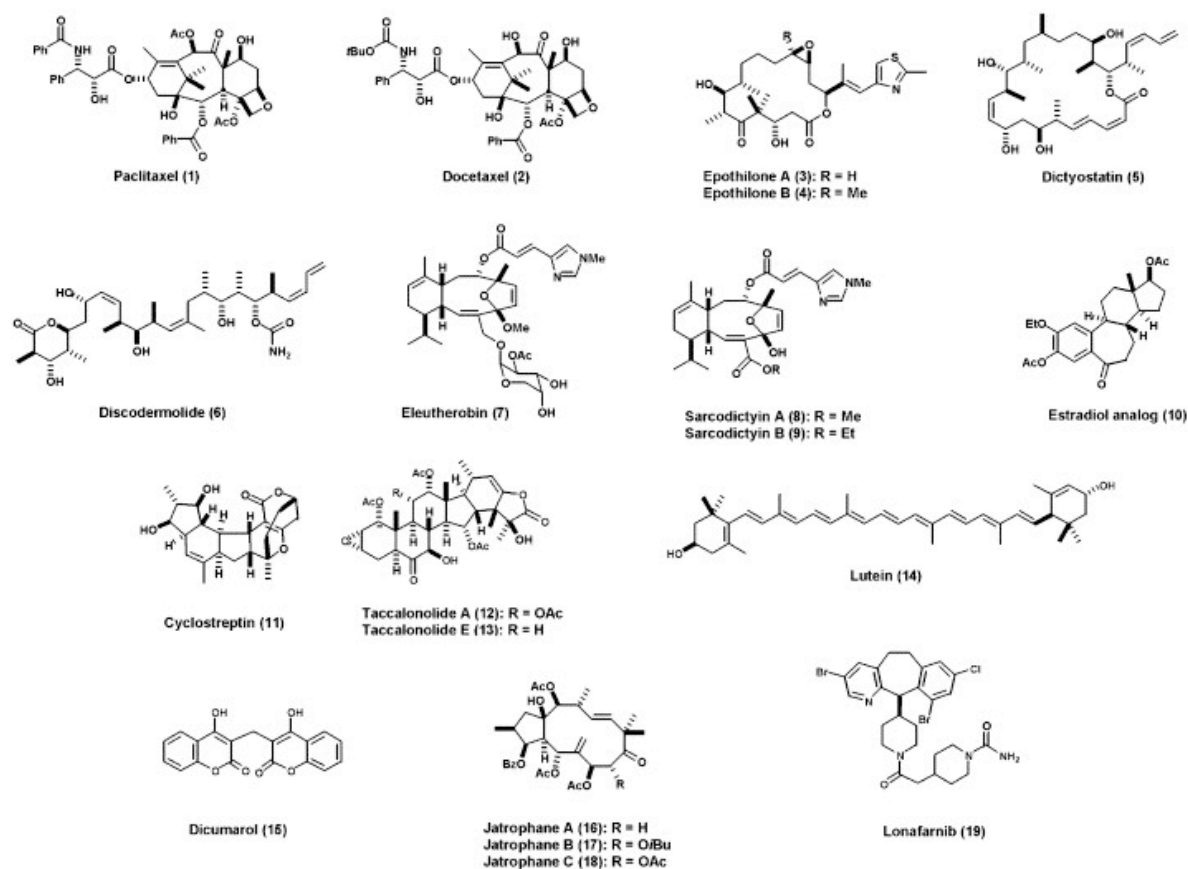


**Figure 2.2:** Interaction of diverse tubulin-binding molecules (vinblastine, colchicine and paclitaxel) at diverse binding sites on microtubules. (Figure from Jordan et al.<sup>43</sup>)

Both mechanisms ultimately interfere with the precise timing of the cell cycle and hamper the cell's ability to structure its cytoskeleton in a flexible manner, thus thwarting cell regulation and growth. The resulting defective cell usually then undergoes apoptosis or cell death. Since cancer cells divide more rapidly than normal ones, such drugs block their division and prevent their growth.

In the last few decades, several molecules interfering with microtubule assembly have become part of the armamentarium of cancer therapy. Taxol is the most famous example. An epothilone analog called ixabepilone was approved by the FDA for use against aggressive breast cancer<sup>49</sup>. Part of the impetus in the discovery of microtubule-binding

molecules is the sheer variety of structures that play this role<sup>50</sup>, an impression rendered clear by the very different looking structures of tubulin-binding molecules below (**Figure 2.3**).

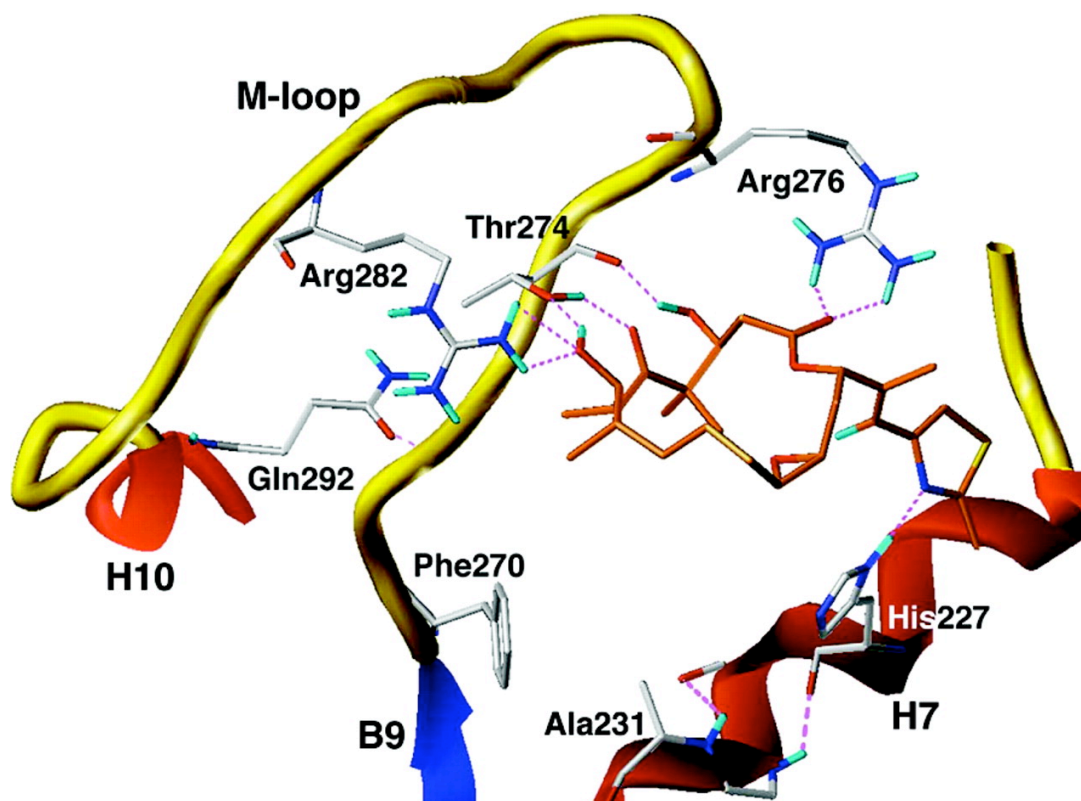


**Figure 2.3:** Molecules binding to tubulin (Figure from Buey at al.<sup>50</sup>)

These molecules also have diverse binding characteristics. For instance, taxol, epothilone, discodermolide and dictyostatin are thought to bind to the same site at the interface of the alpha and beta tubulin dimer whereas laulimalide and peloruside are thought to bind to a different site<sup>51</sup>. This fact also underscores the value of these molecules as potential

anticancer agents; since they bind at different sites and exercise different actions, they could purportedly have efficacy against resistant tumors wherein one drug is rendered inefficient by mutations that block its binding to a specific site. The drugs also result in variable microtubule morphologies suggesting they modulate microtubule assembly or disassembly through complex and subtly different mechanisms<sup>52</sup>, a fact that can also be potentially exploited in multipronged combination therapy.

Our group has had a long-standing and continuing interest in tubulin and tubulin-binding drugs. Through several collaborative computational and experimental studies, the group has elucidated the bioactive conformations of taxol and epothilones in the tubulin binding site by making use of electron density from electron crystallography. In case of taxol, a proposed novel conformation called T-taxol was utilized to construct hypotheses about constrained analogs that would show improved potency<sup>53-55</sup>. These hypotheses were validated through several rounds of synthesis and biological testing and were consistent with the SAR data for the compounds<sup>53</sup>. Similar SAR data combined with photoaffinity labeling and site-directed mutagenesis was utilized for constructing a model for epothilone in the tubulin binding site<sup>33</sup> (**Figure 2.4**).

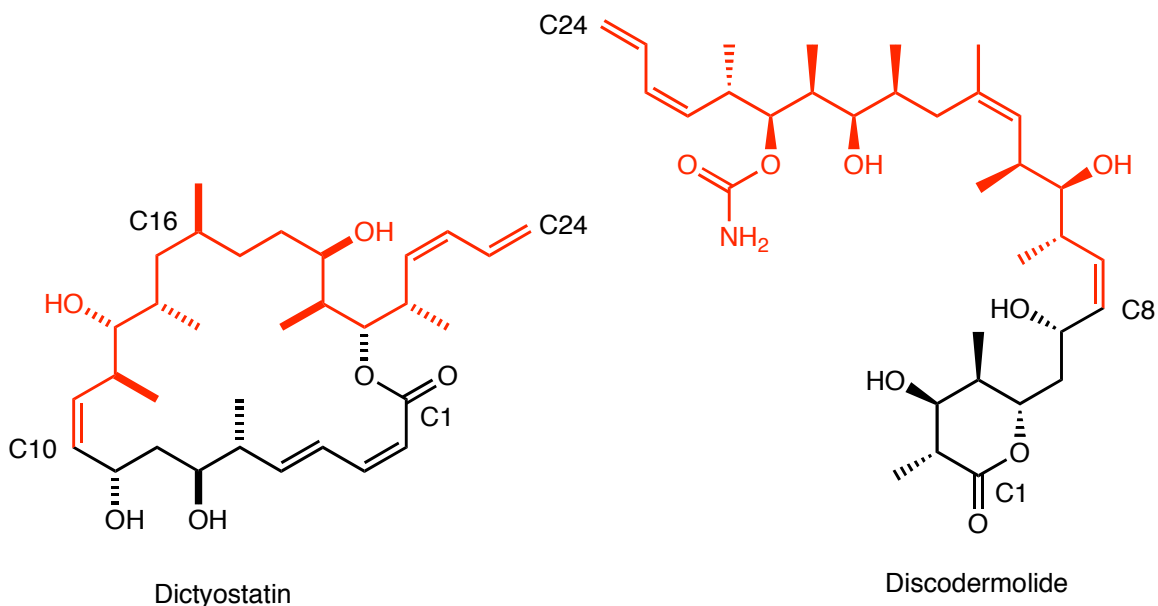


**Figure 2.4:** Binding mode of epothilone A in tubulin (Figure from Nettles et al.<sup>33</sup>)

Preceding the binding mode analysis, NAMFIS analysis of both epothilone and taxol showed that the bioactive conformations for the molecules exist as small percentages (2-5%) in solution. NAMFIS analysis was also used to derive solution populations for laulimalide although the exact binding site of laulimalide is not known. However, it is of continuing interest in this group to investigate solution as well as bound conformational profiles of other tubulin-binding agents.

## 2.2: Dictyostatin and Discodermolide

Our interest was particularly engaged by two recent extremely potent microtubule-binding agents that have been studied extensively; discodermolide (DCT) and dictyostatin (DDM) (Figure 2.5)



**Figure 2.5:** Dictyostatin (DCT) and discodermolide (DDM), with the red parts indicating similar substitution patterns.

Both DCT and DDM are marine macrolides isolated from sponges. DDM was isolated in 1990 from the sponge *Discodermia dissoluta* by Gunasekera and others at the Harbor Branch Oceanographic Institute<sup>56</sup>. Its complete structural characterization had to await its total synthesis by Schreiber et al.<sup>57</sup> DDM was found to be a highly potent promoter of microtubule assembly<sup>58</sup> and most importantly, retained potency against cell lines that had mutations rendering them resistant to PTX. For example, Kowalski et al. showed in 1997 that resistance to DDM in the 1A9 cell line bearing mutations in  $\beta$ -tubulin is essentially absent compared to 25-fold resistance to PTX in the same cell line<sup>59</sup>. These researchers

also demonstrated that DDM resistance to the SW620AD colon carcinoma cell line overexpressing the P-glycoprotein transporter protein is only 25 fold relative to the wild type, compared with 930 fold for taxol. Thus, DDM appears to escape the two dominant mechanisms in cells that confer resistance to taxol; the emergence of mutations in the binding site and the overexpression of efflux pumps that reduce the concentration of the drug inside the cells. In addition, action of DDM against microtubules also demonstrates some special consequences such as induction of accelerated cell senescence<sup>60</sup>.

Microtubules polymerized in the presence of DDM are also more stable compared to those with paclitaxel. Competition experiments with paclitaxel indicated discodermolide to bind to the paclitaxel site<sup>58,61</sup>. DDM also showed substantial activity as an immunosuppressant. Interest in DDM remains high because in spite of further discovery of several other microtubule-binding agents, DDM remains the most potent promoter of microtubule assembly known to date. While DDM was recently withdrawn from clinical trials because of hepatotoxicity, the search for potent and non-toxic analogs continues<sup>45</sup>.

Dictyostatin was isolated from a sponge of the genus *Spongia* in the Maldives and was found to be an extremely potent antiproliferative agent with nanomolar potency<sup>62,63</sup>. Like discodermolide, it was found to be a highly potent promoter of microtubule assembly. It also retains substantial activity against taxol-resistant cell lines<sup>62</sup> (**Table 2.1**).

Table 1: 50% Growth Inhibitory Concentrations (GI<sub>50</sub>) of Dictyostatin, Discodermolide, and Paclitaxel against Human Ovarian Cancer Cell Lines after 72 h in Continuous Presence of the Agents as Determined by the MTS Assay<sup>a</sup>

test agent	GI <sub>50</sub> ± SD, nM (fold-resistance)		
	1A9	1A9PTX10	1A9PTX22
dictyostatin	0.69 ± 0.80	3.2 ± 2.4 (4.6)	1.3 ± 1.0 (1.9)
discodermolide	1.7 ± 1.2	6.2 ± 3.6 (3.6)	7.0 ± 8.4 (4.1)
paclitaxel	0.71 ± 0.11	64 ± 8 (90)	51 ± 9 (72)

<sup>a</sup> 1A9 cells express wild-type  $\beta$ -tubulin, whereas the 1A9PTX10 and 1A9PTX22 cells express, respectively, the Phe270→Val and Ala364→Thr mutant forms of the protein.

**Table 2.1:** GI<sub>50</sub> values for taxol, DDM and DCT inhibition of wild-type and mutant tubulin strains. The table indicates that while the potency of taxol for the PTX10 and PTX22 cell lines is attenuated, DCT and DDM are almost equally efficacious against them compared to wild-type tubulin (Table from Madiraju et al.<sup>62</sup>)

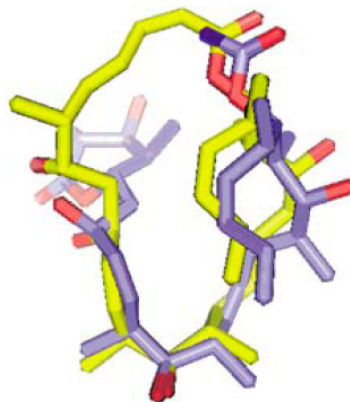
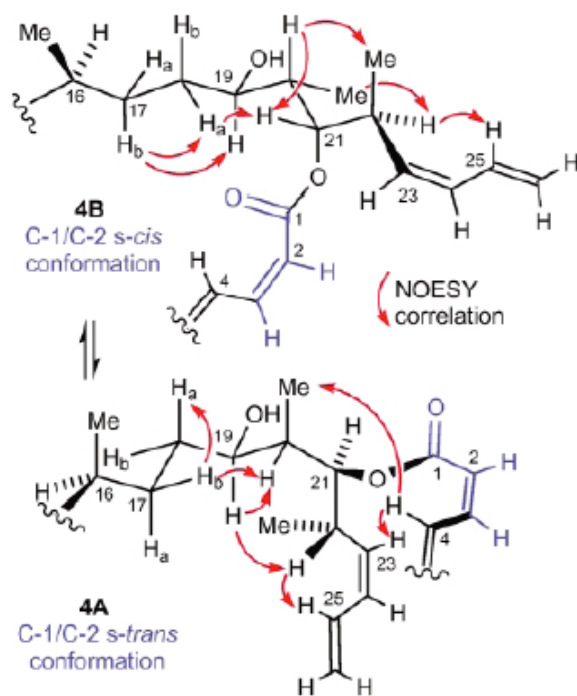
Even before its complete configurational assignment was achieved with spectroscopic techniques and total syntheses<sup>64-66</sup>, the structural resemblance between DCT and DDM was apparent. The two molecules especially share the substitution patterns of methyl and hydroxyl functionalities in much of their structure; C8-C24 for DDM and C10-C24 for DCT (Figure 3). However, minor differences in these structurally similar sections are also reflected in the SAR data for the two molecules. For instance, there is no 16-Me present in DDM whereas the corresponding 16-Me in DCT is thought to have a productive interaction with a hydrophobic phenylalanine in the tubulin binding site<sup>67</sup>. DDM and DCT also induce similar changes in microtubule assembly and have comparable potencies<sup>62</sup>. In the light of these similarities, several hybrid DDM-DCT constructs have been synthesized and they have been shown to be potent microtubule stabilizers<sup>68,69</sup>.



However, in spite of the structural resemblance between DDM and DCT, hypotheses concerning common binding modes for the two molecules have to be carefully formulated. Protein binding sites are diverse and can potentially bind ligands in many dissimilar binding orientations. Also, observations of unexpected and dramatic changes in binding modes and SAR resulting from small changes in ligand structure have a long history in medicinal chemistry<sup>13</sup>. Often the addition of a single hydrogen bonding functionality is sufficient to switch the binding mode in order to take advantage of the new hydrogen bond. Thus, one has to be circumspect in postulating similar binding modes even for molecules that differ in their structure by a single chemical functionality; when the binding involves molecules like DDM and DCT which demonstrate many noticeable structural differences, it would certainly be prudent to postulate similar binding motifs only after a careful and comprehensive analysis of the existing data.

### 2.3: *Dictyostatin*

It was such a hypotheses about similar binding modes for DCT and DDM that prompted our interest, first in dictyostatin. In 2004, Paterson et al. published a detailed assignment of the configuration of DCT by NMR experiments, including 2D NOESY spectroscopy<sup>66</sup>. In this study the authors also proposed a dominant set of conformations for DCT based on characteristic resonances in the NMR spectra. More specifically, the C1-C16 section of DCT was proposed to be “relatively rigid” while the C16-C24 section was proposed as converting between two conformations, one in which the lactone was pointed “in” (*s-cis* configuration) and the other in which it was pointed “out” (*s-trans* configuration). The latter conclusion was based on the observation of resonances for both *cis* and *trans* conformations (**Figure 2.6, a**).



**Figure 2.6. a.** Proposed set of conformations (*s-cis* and *s-trans*) for the C16-C24 segment of DCT based on NMR data, specifically NOESY correlations (Figure from Paterson et al.<sup>66</sup>) **b.** Proposed overlap of dominant conformation of DCT with x-ray conformation of DDM (Figure from Paterson et al.<sup>65</sup>)

Finally, the study postulated that a global minimum conformer from a MM2\* conformational search provided an ‘adequate fit’ to the NMR data for the *s-trans* conformer. This conclusion was based on a comparison of calculated and observed  $^3J_{\text{H-H}}$  coupling constants and based on the presumed fit of this data, the MM2\* global minimum was proposed as the dominant conformation for DCT in methanol solution.

The Paterson et al. study caught our attention for two reasons. Firstly, the discussion about multiple interconverting conformations of molecules in the introduction makes it clear that it is unlikely for a flexible molecule such as DCT with 21 rotatable bonds to exist as a single dominant conformation with limited conformational variability only between the C16-C24 side chain. Secondly, in this report and another one<sup>65</sup>, the dominant conformation deduced for DDM was also overlaid with a conformation for DDM that had been mapped by transfer-NMR experiments (**Figure 2.6, b**). The superposition of the two molecules was rather arbitrary and seemed to be accomplished largely by visual inspection. Also as stated above, the conformation of DCT chosen for overlap with DDM was a single conformation satisfying the average NMR data. To further investigate the conformational space explored by DCT in solution, we decided to perform a NAMFIS analysis of DCT using the NMR data provided in the original report<sup>70</sup>.

The work was initiated by examining the proposition<sup>66</sup> that dictyostatin’s experimental coupling constants and those calculated from the MM2\* global minimum correspond to similar structures. Surprisingly, comparison of the list<sup>66</sup> of experimental and calculated coupling constants indicates that there are significant discrepancies between several pairs

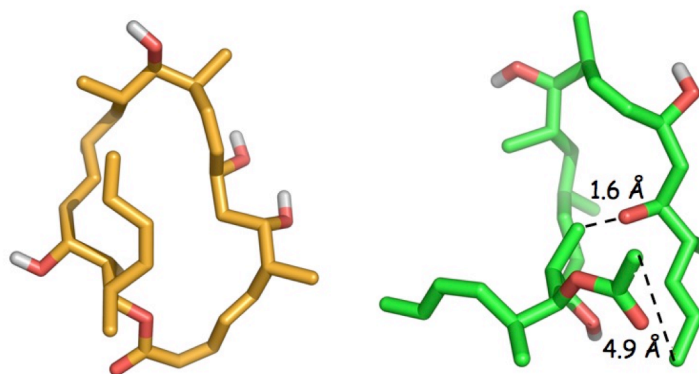
of coupling constants, which in turn are expected to translate into substantial differences between the dihedral angles derived for the two structures. For example, the  $^3J_{\text{H-H}}$  values for H5-H6 and H8b-H9 differ by 4.4 and 8.7 Hz, respectively. The coupling constants for H18a-H19, H19-H20, and H20-H21 are mismatched by 5.9, 4.3 and 5.0 Hz respectively. The Karplus equation indicates that these differences in coupling constants will translate into significant differences in the corresponding dihedral angles. A full list of the  $^3J_{\text{H-H}}$  values and dihedral angles corresponding to the calculated and experimental coupling constants is provided in **Table 2.2**. It is noteworthy that, between the two sets of data, 6 torsional angles differ by at least 10-20°, while 12 differ by greater than 20°. In order to confirm the parameters calculated from the MM2\* global minimum by Paterson et al., we duplicated the MM2\* conformational search and recalculated the coupling constants for the GM structure. These values are in column 2 of **Table 2.2**

H-H Fragment	Calc-1 <sup>a</sup>	Calc-2 <sup>b</sup>	Expt	$\varphi(\text{HC-CH})^c$	$\varphi(\text{HC-CH})$	$\varphi(\text{HC-CH})$
	$^3J_{\text{H-H}}$ (Hz)	$^3J_{\text{H-H}}$ (Hz)	$^3J_{\text{H-H}}$ (Hz)	<b>2</b> , deg	<b>3</b> , <sup>d</sup> deg	$ \text{D}(\mathbf{2} - \mathbf{3}) $ , deg
H5 to H6	11.1	11.8	6.7	174.9	135.7	<b>39.2</b>
H6 to H7	1.8	0.3	4.0	59.3	134.5	<b>75.2</b>
H7 to H8a	11.3	9.7	10.6	164.6	154.0	10.6
H7 to H8b	1.0	0.9	2.7	79.7 (-78.8)	-53.0	<b>25.8</b>
H8a to H9	4.0	6.0	3.3	46.2	67.0	<b>20.8</b>
H8b to H9	1.4	0.9	10.1	68.9 (-66.5)	-152.0	<b>85.1</b>
H9 to H10	5.6	7.0	9.5	129.0	144.8	15.8
H10 to H11	9.4	12.2	11.1	0.3	NA	-
H11 to H12	11.2	11.9	11.1	178.6	164.8	13.8
H12 to H13	1.2	0.2	3.1	65.6 (-67.1)	-38.6	<b>28.5</b>
H13 to H14	9.8	10.3	8.0	167.6	138.2	<b>29.4</b>
H14 to H15a	3.0	3.2	3.8	61.1 (-61.9)	-56.4	5.5
H14 to H15b	12.3	11.9	11.2	176.4 (-176.1)	-163.1	13.0
H15a to H16	12.2	11.9	10.3	174.3 (-175.0)	-155.5	19.5
H15b to H16	3.2	3.3	3.8	60.2 (-61.8)	-56.7	5.1
H16 to H17a	3.0	3.1	Sm	61.7 (-64.8)	NA	-
H16 to H17b	12.3	11.9	9.0	178.5 (-173.7)	-146.8	<b>26.8</b>
H17a to H18a	12.8	14.0	12.8	172.8 (-173.0)	-160.6	12.4
H17a to H18b	1.6	1.3	Sm	71.3	NA	-
H17b to H18a	3.7	3.8	4.7	56.4 (-58.3)	-51.4	6.9
H17b to H18b	12.8	14.0	12.8	172.2	160.9	11.3
H18a to H19	11.7	10.6	5.8	170.4 (-170.7)	-138.1	<b>32.6</b>
H18b to H19	2.5	5.1	2.0	55.0 (-55.9)	-78.0	<b>22.1</b>
H19 to H20	1.5	1.4	5.8	61.6 (-64.8)	-27.9	<b>36.9</b>
H20 to H21	9.9	9.3	4.9	171.1	137.8	<b>33.3</b>

**Table 2.2:** Dictyostatin  $^3J_{\text{H-H}}$  values, the corresponding dihedral angles and the difference in torsion space between the MM2\* GM structure (**2**) and the experimental structure (**3**).

- <sup>a</sup> Calculated by Paterson et al. from the dihedral angles of the MM2\* modeled dictyostatin global minimum using the modified Karplus equation of Haasnoot, de Leeuw, and Altona.
- <sup>b</sup> Calculated for the MM2\* GM from the duplicated conformational search using the modified Karplus equation of Haasnoot, de Leeuw, and Altona incorporated in Macromodel (v. 7.2).
- <sup>c</sup> Negative-signed parenthetical values back-calculated from the MM2\* GM.
- <sup>d</sup> Back-calculated from the experimental  $^3J_{\text{H-H}}$  values and double-checked with the modified Karplus equation incorporated in Macromodel (v. 7.2).
- <sup>e</sup> In-house software written at Emory University: Zhong, S.; Jogalekar, A. S.; Snyder, J. P., unpublished.

A comparison of the structure calculated from the experimental coupling constants and the MM2\* GM purported to be similar to it indicates the dramatic differences between the two. **Figure 2.7** illustrates that **3** is acyclic with a 4.9 Å gap between carbons that require a 1.4 Å union in order to create a closed dictyostatin ring. In addition, a 1.6 Å steric clash between the 17-OH oxygen and the C-20 methyl carbon underscores the high-energy nature of this virtual structure. Thus the comparison reveals that structures **2** and **3** are quite different in their overall geometries. While **3** is clearly a virtual structure, we propose below that conformer **2** is most likely in the same category. As has been noted in previous studies, the assignment of averaged NMR data to a single conformation inevitably leads to a high-energy virtual structure<sup>71</sup>.



**Figure 2.7:** Comparison of the Paterson et al. MM2\* global minimum **2** (gold) and **3** (green), a structure derived from the  $^3J_{\text{H-H}}$ -derived dihedral angles. For **3**, a 4.9 Å gap separates two bound carbons in **2**. Structure **3** also sustains a 1.6 Å cross-molecule steric clash between a hydroxyl oxygen and a methyl carbon (cf. **Table 2.2**).

#### **2.4: Dictyostatin NAMFIS conformations**

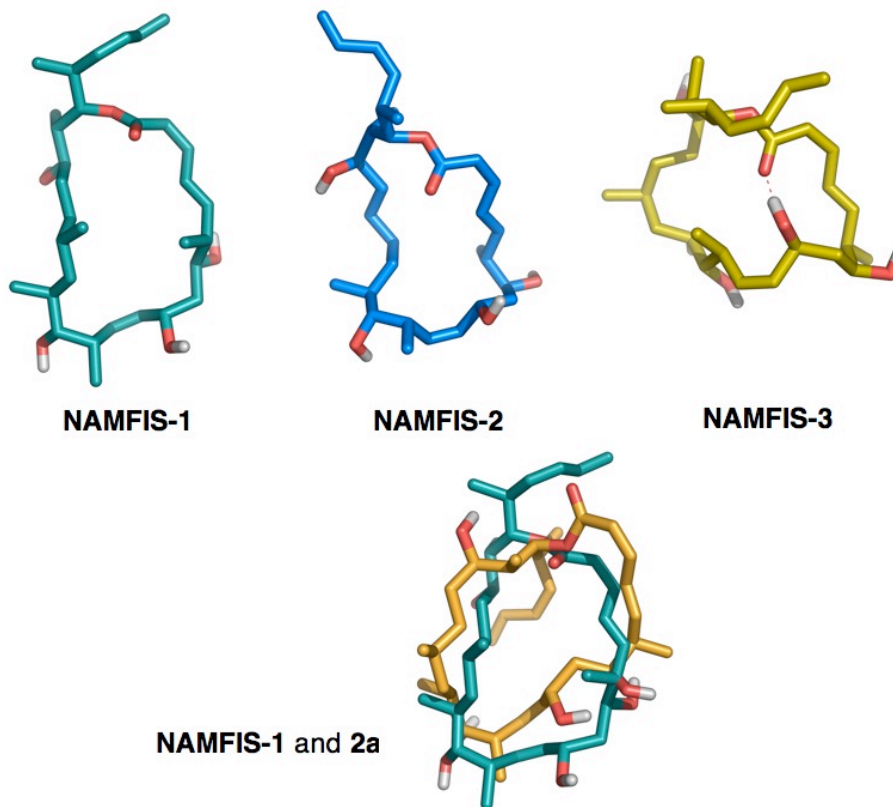
To generate a meaningful conformational pool for DCT, conformational searching was carried out with three force fields (MMFF, AMBER\*, and OPLS-2005) using the mixed low-mode/Monte Carlo method and the GBSA/H<sub>2</sub>O implicit solvation model in Macromodel. An energy cutoff of 10 kcal/mol was used to ensure extensive adequate conformational coverage. In each of the three searches, the global minimum was found at least 10 times, indicating that the torsional energy surface has been exhaustively sampled. The conformations from each force field were further minimized to convergence with the same force field by subjecting them to 50 steps of Full Matrix Newton-Raphson optimization. Conformers from the three searches were pooled and



duplicates discarded. The resulting collection of 2053 distinct conformations served as input for the NAMFIS study.

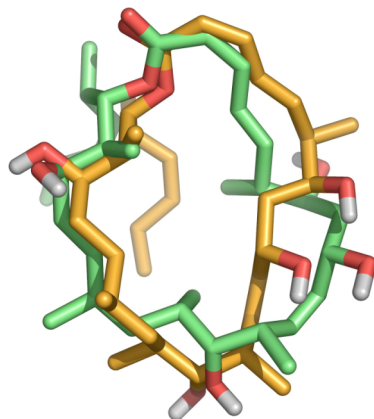
NAMFIS analysis was performed with the Cambridge-HOI NMR dataset and the 2053 unique optimized conformations described in the Methods section to provide a pool of 16 conformers with estimated populations ranging from 11 to 2%. Of these the first (11%) and fourth (8%) differ only by an OH rotation, contributing a total of 19% to the conformer pool.

Given that the NMR data is not quantitative (i.e. NOESY cross-peaks were catalogued as very strong to weak), this data-to-structure fit achieves an SSD of 150. Its significance, however, is highlighted by the fact that **2** constructed with the same data delivers an SSD of 1733. The NAMFIS ensemble of conformations is clearly a superior solution to the data-fitting problem. Seven of the top sixteen conformations, including the top three (**Figure 2.8**), display the *s-cis* lactone-in conformation (52% of the total ensemble). In this respect, the NAMFIS analysis agrees qualitatively with the Paterson et al. deduction<sup>72</sup> that there is an approximately even distribution of *s-cis* and *s-trans* forms.



**Figure 2.8:** Top three NAMFIS conformers: best fit (NAMFIS-1, 11%), NAMFIS-2 (11%) and NAMFIS-3 (9%). The latter is predicted to enjoy an internal hydrogen bond. The pair below superposes **2** (orange) with NAMFIS-1; all-atom RMSD 4 Å.

This observation is at odds with the proposal that *s-trans* lactone-out **2** is the dominant conformation in solution. The NAMFIS conformation closest to **2** in terms of RMSD is the 6<sup>th</sup> best fit present to the extent of 7%. However, even this conformation differs substantially from the **2** conformer in several parts of the molecule. **Figure 2.9** shows the overlap of the two conformations with an all-atom RMSD of 2.7 Å.



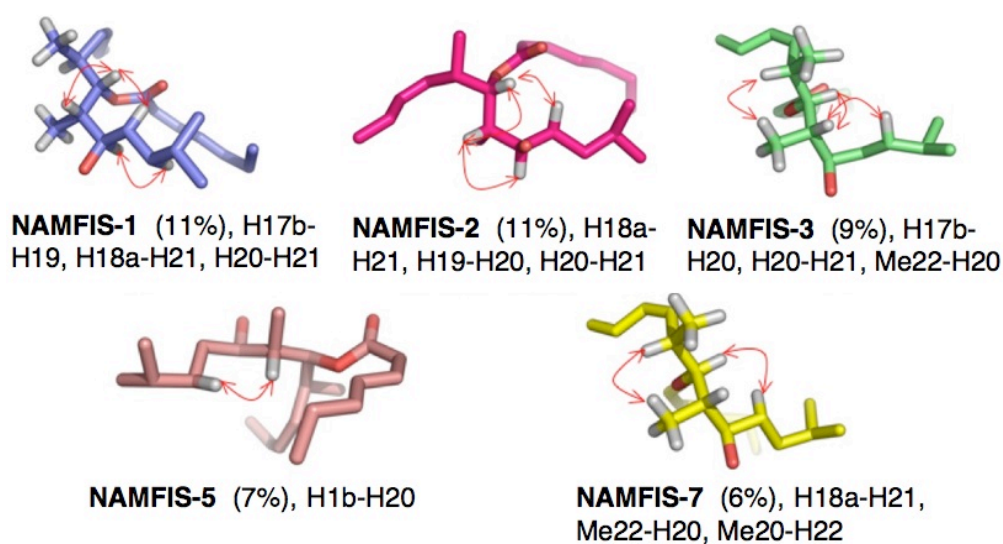
**Figure 2.9:** Overlap of **2** (orange) and the most similar NAMFIS conformation (green, 6<sup>th</sup> best fit, 7%); RMSD 2.7 Å.

As noted above, the Cambridge-HOI team not only deduced the full stereogenic assignment for DCT, but also proposed the existence of at least two rapidly equilibrating conformations in the C16-C23 region. The *s-trans*, lactone-out form (**2a**) was proposed to be more stable than the *s-cis*, lactone-in (**2b**) conformer based on relative MM2\* energies. The latter relative stability prompted the identification of the MM2\* *s-trans* global minimum **2a** as the dominant conformer in solution. Unfortunately, force field energies for structures as complex and polar as **1** are fickle and ordinarily difficult to correlate with experiment especially because of an overemphasis on electrostatic interactions. This is borne out by the fact that force field optimizations for complex molecules with different force fields usually don't agree with respect to the structure of the global minima and almost always display dissimilar energies for even similar minima<sup>32</sup>.

NAMFIS, on the other hand, not only avoids the ambiguities of intuitively disentangling complex NMR spectra but also bypasses the pitfalls of attempts to predict relative energies in solution. It simultaneously eliminates the temptation to arbitrarily select a global minimum from a specific force field as an experimental conformation. By extracting conformer mole fractions in solution based on structure, as is the case with NAMFIS, the relative free energies can be calculated directly from the populations. In the present case, conformer **2** is not among the ensemble of conformers derived from the CD<sub>3</sub>OD data measured for **1** (cf. **Figure 2.9** and SSD comparison). Likewise, the *s-cis*, lactone-in motif is the favored geometry posited for dictyostatin in the same solvent.

It is pertinent to ask whether the ensemble of dictyostatin conformers obtained by the NAMFIS analysis accounts for all the NMR parameters provided in the Cambridge/HOI report. This is a stringent criterion for evaluating NAMFIS performance, because if any of the NMR data is not satisfied by one or more of the NAMFIS conformers, then it is likely that complete coverage of conformational space has not been achieved. Gratifyingly, the top sixteen NAMFIS conformers collectively fit every piece of NMR data that is observed for both the rigid as well as the flexible portions of dictyostatin. At the same time, no single conformer satisfies all the data. Significantly, for the flexible portion of the molecule (C16-C23), we find that half of the top conformers satisfy the NMR data for the *s-cis*, lactone-in conformation in contrast to the single **2a** *s-trans* conformer arising from MM2\* energies.

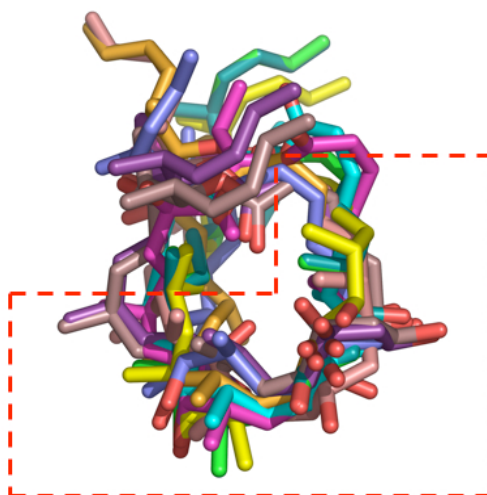
**Figure 2.10** displays the flexible C16-C23 portions of the molecule for some of the top NAMFIS conformations and the NOE data that is matched. The remainder of the molecule, C2-C16, is proposed to be relatively rigid<sup>72</sup>. Evaluation of the NAMFIS conformers indicates that the top conformer contributes strongly to the NMR data in the C10-C16 region, but like the remaining fifteen members of the NAMFIS pool, accounts for structural features within C2-C16 and other sectors to limited and various extents.



**Figure 2.10:** C16-C23 fragments for some of the top NAMFIS conformers and the partial NOE data which they satisfy.

The conclusion that dictyostatin is relatively rigid in the C2-C16 fragment would be premature. There are many conformations associated with the C2-C16 fragment that can contribute to a fit of the  $\text{NOE}/^3J_{\text{H-H}}$  data. **Figure 2.11**, superposing C2-C16 for the second to tenth most populated NAMFIS structures on the most populated NAMFIS-1,

makes it clear that this segment of **1** cannot be regarded as even relatively rigid. The molecule is characterized by a spread of conformations. RMSDs for ring carbons of NAMFIS-2 to NAMFIS-10 relative to NAMFIS-1 in the C2-16 region range from 0.2 Å for the 4th ranked conformer (virtually identical) to 1.8 Å for the 10th ranked conformation.



**Figure 2.11:** Superposition by ring atoms in the C2-C16 region for the most populated ten NAMFIS conformers. RMSDs relative to the top ranked conformer (NAMFIS-1) range from 0.2 to 1.8 Å. The red outline encompasses the C2-C16 segment.

Thus, all sixteen NAMFIS conformers contribute intensity to the NOE cross-peaks and  $^3J_{\text{H-H}}$  values that suggest relative rigidity when analyzed as one or two conformers.

### ***2.5: NAMFIS conformations in DMSO- $d_6$ and CD $_3$ OD***

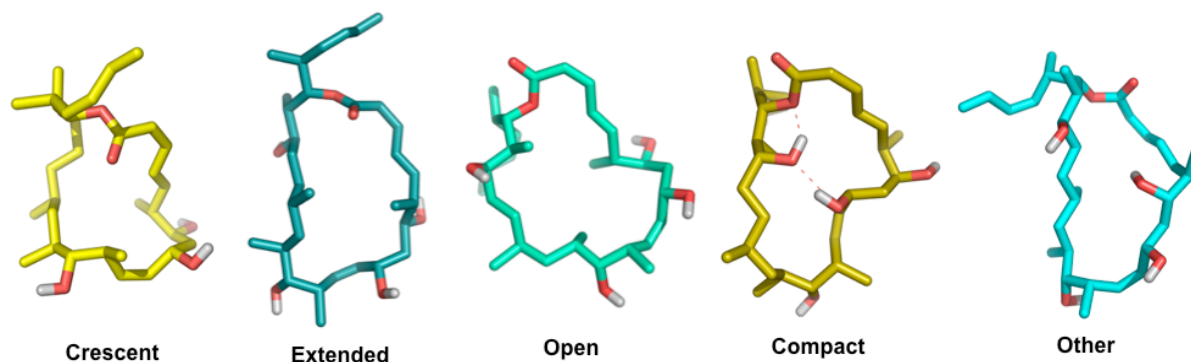
To perform a conformational analysis of dictyostatin based on more quantitative data and to investigate solvent effects on conformational preferences, we decided to apply NAMFIS to data for dictyostatin obtained in methanol. To achieve this we secured the collaboration of Prof. Dennis Curran and Drs. Damodaran Krishnan and Won-Hyuk Jung of the University of Pittsburgh. Prof. Curran's group was one of the first two groups to achieve the total synthesis of DCT and had a sample of DCT for NMR studies.

All NMR experiments in DMSO were performed on a Bruker Avance spectrometer operating at 600 MHz and equipped with a XYZ-gradient triple resonance probe. Spectra were processed using Topspin and analyzed using the iNMR software to extract interproton distances. The sample was prepared by dissolving 1.7 mg of Dictyostatin in 0.5 mL of DMSO- $d_6$ .  $^1\text{H}$  and all 2D spectra were accumulated at 298K. Five ROESY spectra were recorded at 70, 100, 150, 180 and 200ms mixing times to check the linearity of the cross-relaxation buildup rate. The acquisition times  $t_1$  and  $t_2$  for the ROESY experiments were 207 and 26 ms respectively. Relaxation delay was set to 3s and 80 scans were accumulated per  $t_1$  increment.

Thus, the Pittsburgh scientists extracted 35 NOE distances and 15  $^3J_{\text{H-H}}$  values from the averaged NMR spectrum. When processed by NAMFIS, this data gave 15 conformations ranging from 1-20%. The previous two MM2\* conformations are again absent from the conformational ensemble.

The NAMFIS analysis reveals that out of the 15 conformations, only 3 have the lactone *s-trans* while all others including the top 3 have the lactone *s-cis*. Thus the *s-cis* form seems to be even more dominant in DMSO than it is in methanol. Intramolecular hydrogen bonding among the conformers is also consistent with the nature of the solvent; 46% conformers are hydrogen-bonded in DMSO while only 18% are hydrogen-bonded in methanol, which is in agreement with the fact that DMSO can only accept hydrogen bonds while methanol can act both as a donor and acceptor.

Clustering of the conformations by inspection allows us to classify them into five dominant families as illustrated in **Figure 2.12**.



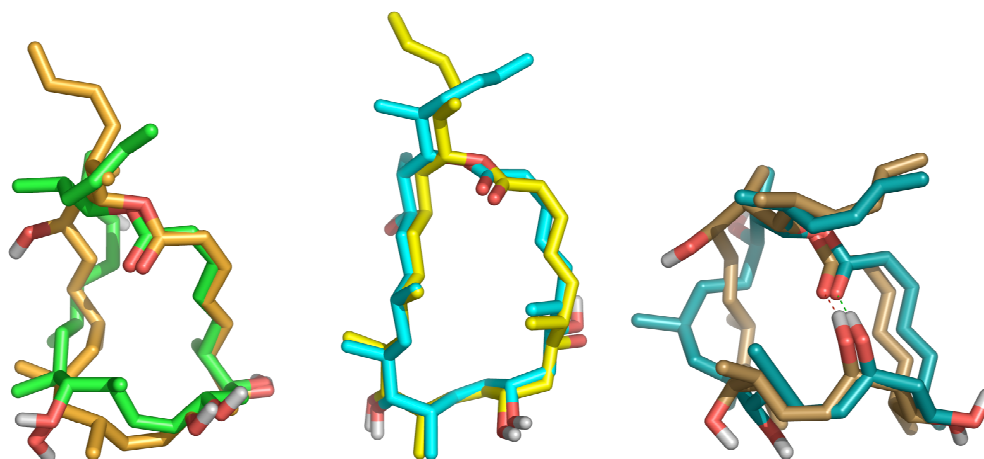
**Figure 2.12:** Dictyostatin families in DMSO

Closer inspection of the conformations reveals common features with those in methanol. The families are named as Crescent, Extended, Open, Compact and Other. Some conformations in the two solvents are very similar, but with differing populations. For example the best-fit DMSO conformation is similar to the 3rd best fit methanol



conformation, while the 3rd best fit DMSO conformation is similar to the best fit methanol conformation. The top DMSO conformation (20%) is identical to the 9th best fit methanol conformation (5%). The similarities between the conformations extend across families. For instance the Crescent family dominates in both DMSO and methanol (44 and 45% respectively) with the Extended family being the second-most dominant one in methanol (23%) and third-most dominant in DMSO (19%). The Extended family is also encompassed among the top 5 conformations in both solvents. The Open family is present only in methanol (14%) as the third-most dominant family while the Compact family is the second-most dominant in DMSO (21%) and fourth-most dominant in methanol (4%). The conformations from the Other families are present to the extent of 13% and 8% in methanol and DMSO respectively and while their features cannot be completely ascribed to any one of the other families, they share partial features from the other family members. The former two MM2\* conformations in methanol, while not found in both solvents, belong to the Compact family.

These similarities are illustrated in **Figure 2.13** and indicate that the solvent change essentially enforces a change in distribution of the conformational families without altering the conformations themselves.



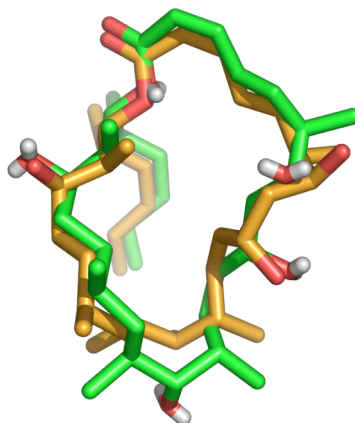
**Figure 2.13:** Superposition of similar conformations in DMSO and methanol: From left- DMSO-1 (20%) and methanol-2 (11%) (RMSD 2.3 Å), DMSO-3 (10%) and methanol-1 (11%) (RMSD 1.6 Å) and DMSO-2 (13%) and methanol-3 (9%) (RMSD 1.8 Å)

### ***2.6: Dictyostatin conformations in the tubulin binding pocket***

Recently there was a report of an NMR-bound structure of dictyostatin deduced through transfer-NOESY experiments<sup>73</sup>. The conformation deduced was seen to belong to the same family as the previous dominant MM2\* conformation, although it differs significantly from this conformation in its details.

To shed possible light on the differences between the bound and solution conformations, we decided to compare the bound conformation with the above solution conformations by applying NAMFIS. For this purpose, we used the NOE distances for the bound conformation as input for the NAMFIS routine and investigated the fit of each individual solution NAMFIS conformation against this bound data. As a starting point for the

comparison, we investigated the fit of the bound conformation against its own bound NMR data and found that it fit this data with a SSD of 127. Similar processing of each individual solution NAMFIS conformation from both DMSO and methanol against the bound NMR data indicated poor fits with SSDs exceeding 400 for all conformations except one. This was the 7th best fit conformation in DMSO present to the extent of 7% that gave a SSD of 122 when fit against the NMR data for the bound conformation. This conformation belongs to the same family as the bound conformation although it differs from the bound conformation, especially in the C6-C18 region as shown below in **Figure 2.14 a**.

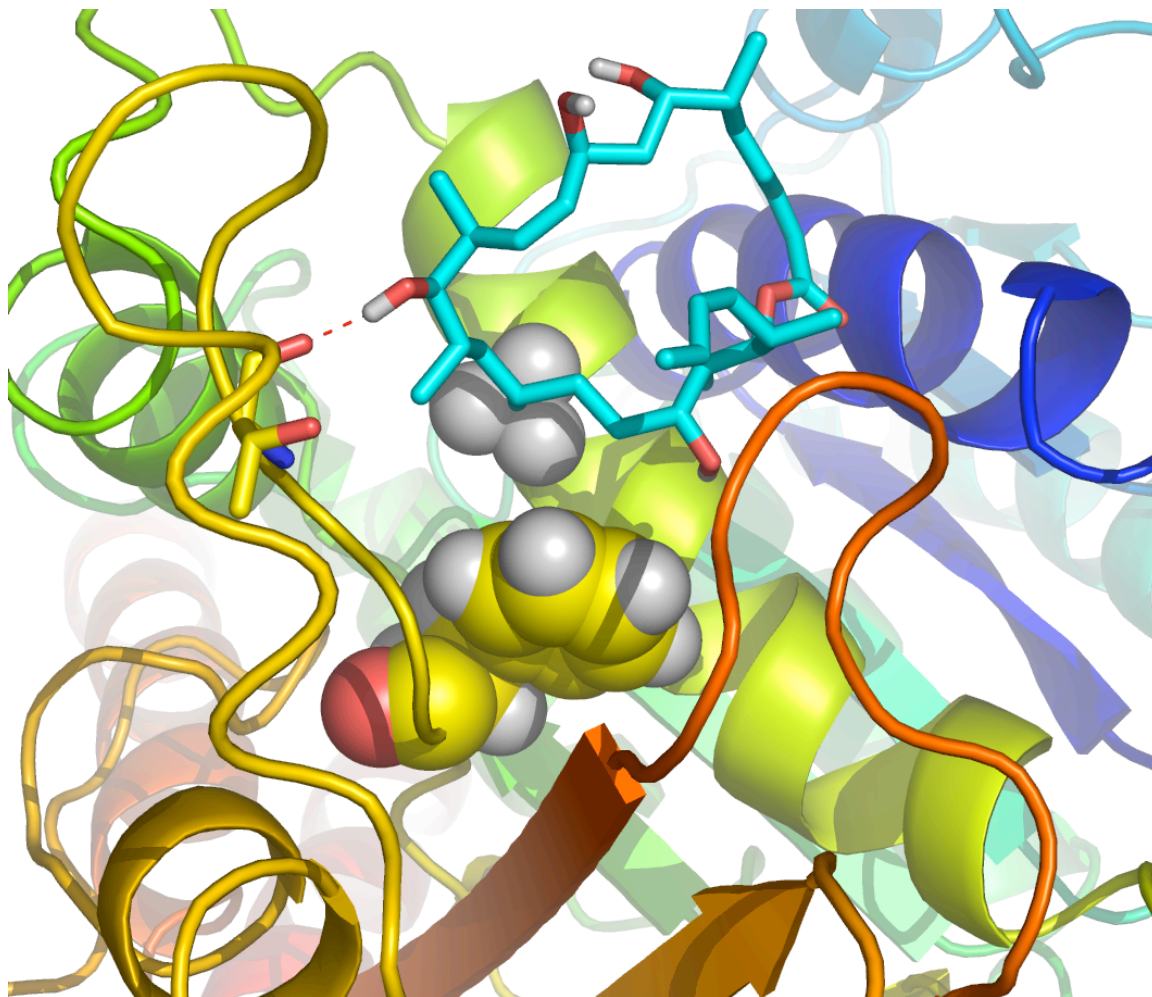


**Figure 2.14 a:** Overlap of NMR bound conformation of dictyostatin and DMSO-7 (7%)

The RMSD is 1.6 Å

Previously in the Canales et al. study the authors had proposed a binding mode for DCT based on an AUTODOCK docking run. However, SAR data was not explicitly taken into account while generating this pose and the pose was rather arbitrarily selected from several possible alternatives. Studies based on photoaffinity labeling<sup>61</sup>, site-directed mutagenesis and synthesis and testing of 16-normethyl DCT analogs<sup>67</sup> have suggested that the 16-Me group on DCT may be in close proximity to Phe 270 in tubulin. To explore this hypothesis we rigidly docked the above proposed DCT conformation from DMSO-d<sub>6</sub> (DMSO-7) into the tubulin structure used previously (1JFF) and generated 20 poses using GLIDE (SP). Since docking scores differed too little to suggest a preference for a particular orientation of the ligand, we used MM-GBSA to try to discriminate between the different poses and generate a more physically realistic orientation. MM-GBSA is a protocol that uses an implicit solvation model (GBSA) to separately calculate the solvation free energies of ligand in solution, protein and ligand in binding pocket. As such it aims to provide a good approximation to relative free energies of binding of similar ligands or conformations and has been used in this capacity before<sup>74</sup>.

Surprisingly, the DMSO-7 conformation was ranked higher than the previously proposed conformation, and the top-ranked MM-GBSA pose places the 16-Me of DCT in relative proximity to Phe 270 of tubulin (Me-C – Phe centroid distance 5 Å) (**Figure 2.14 b**).



**Figure 2.14 b:** Best MM-GBSA binding pose of DMSO 7<sup>th</sup> best fit conformation (7%) in tubulin. The inward looking 16-Me group and Phe 270 in tubulin are rendered as space-filling spheres to depict their proximity. The C13-OH group forms a hydrogen bond with a threonine.

Additionally, all of the other poses differ substantially with respect to the distance between 16-Me and Phe 270 and are ranked lower in free energy than the best pose by a minimum of 1 kcal/mol. Since the pose features a ligand conformation which is a bonafide solution conformation, which is ranked high by the MM-GBSA algorithm and which seems to be consistent with a key piece of SAR data, we propose this pose as an possible alternative model of DCT in the tubulin binding site.

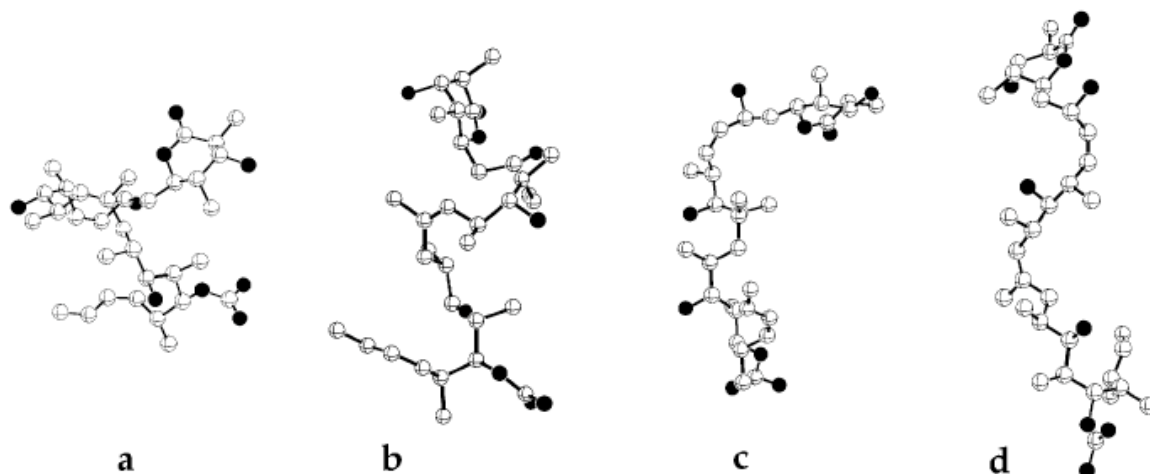
### ***2.7: Outlook and conclusions***

To summarize then, we have performed a comprehensive conformational analysis of dictyostatin in two solvents- DMSO and methanol- using the NAMFIS methodology. Investigation of the conformational profiles indicates common families of conformations in the two solvents with differing populations. The previous proposed pair of dominant methanol conformations are not among them. The study reinforces the pitfalls of deducing single or limited conformations from average NMR data. In addition we have compared the solution conformations with the recently proposed NMR-derived bound dictyostatin conformation and suggest an alternative solution conformation in DMSO that seems to fit the data for the bound conformations equally well. Using docking and MM-GBSA protocols we have derived a binding pose for DCT in tubulin that appears to be consistent with an important piece of SAR data. Future studies based on mutagenesis and analog synthesis and testing will nonetheless test the validity of this and other proposed models and help to shed light on further improvement of this novel potential anticancer drug.

### **2.8: Discodermolide**

After analyzing the conformational profile of DCT we decided to investigate the same for DDM. Unlike DCT, DDM has been crystallized in the solid-state. In addition, a single solution conformation in CD<sub>3</sub>CN for DDM was derived in 2001 by Smith et al. by using information from NMR<sup>23</sup> and was proposed to be essentially identical to the x-ray conformation. However, this conformation was a single conformation derived as a structure from the average NMR data.

A more realistic ensemble of conformations in deuterated DMSO was generated in 2001 by Snyder et al. using coupling constants and distances<sup>75</sup>. The study identified 4 DDM families contributing to the conformational pool. Out of these the ‘corkscrew’ family was the dominant one and seemed to encompass a highly extended version of the x-ray conformation. However, the x-ray structure was significantly different from all members of this family so that it was thought to contribute no more than 1% to the conformational average (**Figure 2.15**). The Snyder et al. study was also performed using conformational ensembles generated with a single force field, namely MMFFs. Since a single force field may be deficient in mapping the conformational surface of a complex molecule, a multiple force field that achieves a conformational deconvolution of DDM would merit consideration. In addition, recent studies have also reported the relevant NMR data for DDM in water, and it would have been instructive to compare DDM populations in the two solvents to explore solvent effects.



**Figure 2.15:** Three families of DDM from the Snyder et al. study in DMSO. a. X-ray structure b. Corkscrew (68%), c. Sickle (21%) and d. Extended (10%)

Recently two groups have also investigated the binding conformation of the drug, most notably by means of transfer NOE displacement experiments with epotholine<sup>76</sup>. The binding conformation derived from these experiments has been conjectured to be very similar to the solid-state conformation<sup>77</sup> and the proposed solution conformation<sup>23</sup>. However, it is the internal conformation of DDM that is in agreement in these two studies, while the binding mode is a hypotheses derived from an AUTODOCK docking run. In addition, this binding mode has not been adequately discussed in the context of the SAR data for the molecule.

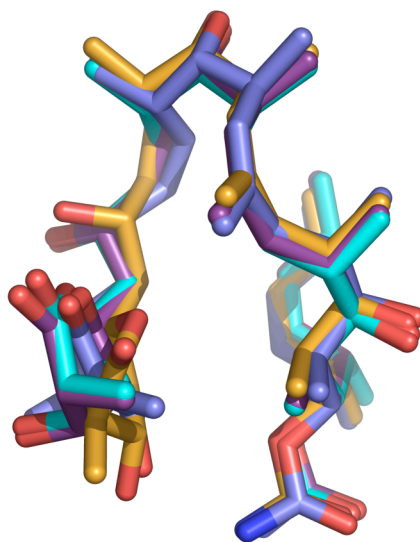
To possibly provide an alternative model for DDM in the tubulin binding site and to further explore its solution conformations in multiple solvents, we have investigated the conformations of DDM by performing NAMFIS<sup>78</sup>. Our objective here was three-fold.



Firstly as stated above, we wanted to improve the previous DDM results in DMSO using a multiple force field search compared to the previous single force field searching protocol. Secondly, using this technique we wanted to investigate whether an alternative conformation of DDM in solution fits the data as well or better than the most recent bound conformation. Thirdly, we wanted to explore a possible alternative conformation of DDM in the tubulin binding site. As was done for DCT, the NAMFIS method was used to deconvolute the averaged NMR spectrum of DDM in DMSO and obtain a Boltzmann population of solution conformations. Surprisingly we find that the conformations determined in the solid-state and proposed in solution are strongly represented in the computationally generated ensembles and NAMFIS results. In the present work, also we examine possible factors influencing the stability of such conformations in various environments. In the previous report<sup>23</sup> by Smith et al., the authors discussed the influence of *syn*-pentane interactions in dictating the solution conformation of DDM. The present study reinforces as well as adds to the previous discussion in three distinct ways by 1) indicating the dominance of these steric interactions in a multiconformational solution ensemble, 2) investigating simplified model systems that underscore the essentially steric nature of the interactions, and 3) using quantum chemical ab initio calculations to rationalize the prevalence of such interactions in the solution and solid-state.

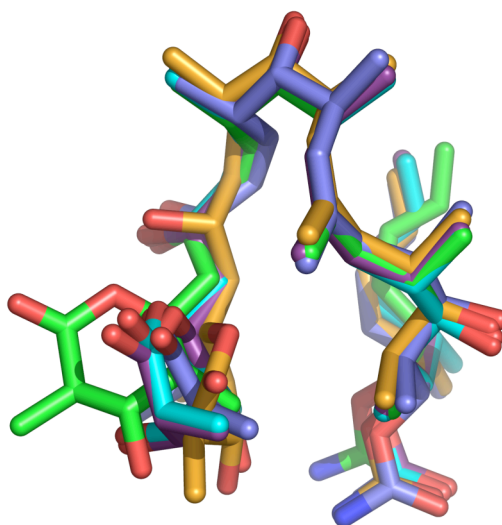
### 2.9: Conformations of DDM in DMSO- $d_6$

Conformational searching and refinement with four force fields (AMBER\*, MMFFs, OPLS 2005 and MM3\*) gave a total of 1282 conformations across the four force fields. Examination of the global minima from the four structure generation methods shows them to be surprisingly similar. This observation is striking in light of the disparate parameterization criteria and charge models used by different molecular mechanics methods to determine energy minima for flexible molecules<sup>32</sup>. **Figure 2.16** superposes the four structures with an average all heavy-atom RMSD of only 0.9 Å. The main variability is in the disposition of the lactone side chain.



**Figure 2.16:** Superposition of the global minima from the four force fields: AMBER\* (blue), MMFFs (cyan), OPLS2005 (gold) and MM3\* (purple). The lactone is at lower left.

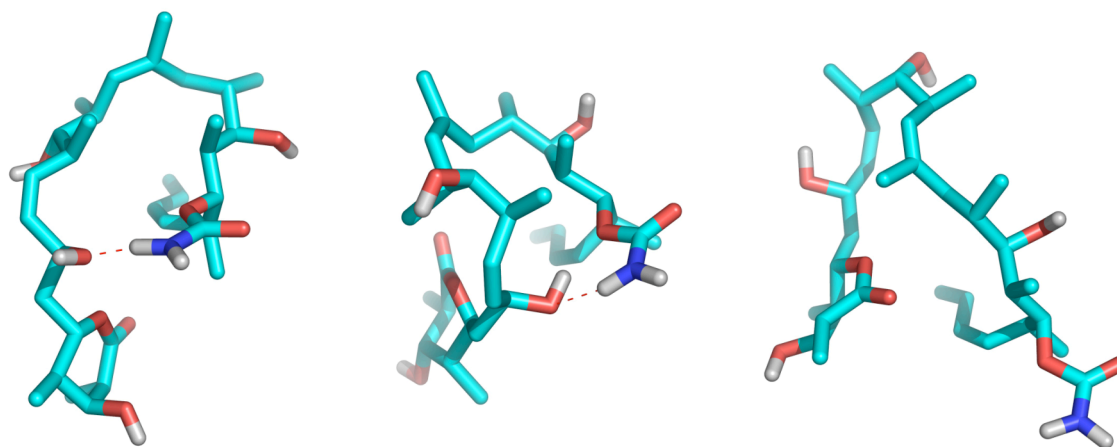
The four global minima are also very similar to the X-ray structure as displayed in the superposition of **Figure 2.17**. The crystal structure (green) presents the lactone in a different orientation compared to the force field structures.



**Figure 2.17:** Superposition of the four force field global minima and the crystal structure (green) of discodermolide.

To complement the force field conformational results with parallel solution studies, discodermolide in DMSO- $d_6$  was examined using NAMFIS. A previous DDM NAMFIS study from this laboratory identified the X-ray conformation as a minor contributor to the solution conformational ensemble<sup>75</sup>. Since that study employed only a single force field, it was repeated using the same NMR data but with a more comprehensive quadruple force field search. As previously, 40 interatomic proton-proton distances from 2D ROESY spectra and 14 three-bond coupling constants ( $^3J_{H-H}$ ) were used for the analysis.

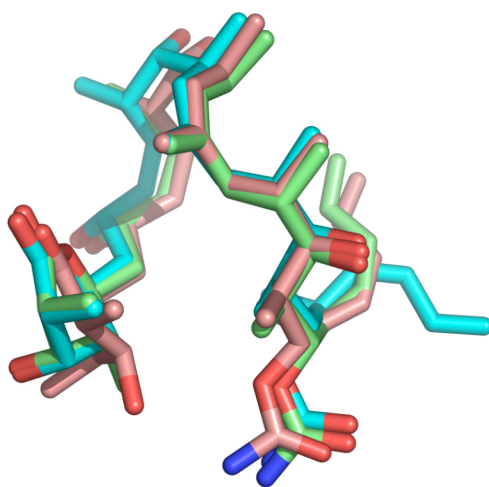
NAMFIS analysis combining the NMR data and the 1282 conformers gave 12 best-fit conformations (NAMFIS 1-NAMFIS 12, SSD = 49) with populations ranging from 2-17%. The significance of this conformational ensemble is highlighted by the fact that the single X-ray conformer alone yields the considerably higher SSD of 246 indicating it to be a rather poor fit to the data by comparison with the 12 conformer pool. The top three NAMFIS conformational families are illustrated in **Figure 2.18**.



**Figure 2.18:** The top 3 NAMFIS conformational families of DDM in DMSO- $d_6$ . From left: Sickle: 17%, Dome: 17% and X-ray: 57% respectively. The top two sustain a hydrogen bond between the carbamate  $NH_2$  and the C-7 hydroxyl. One other conformer belongs to the Extended family (9%)

Inspection of the conformations indicates many defining characteristics. The conformers can be divided into four classes; Sickle (17%), Dome (17%), X-ray (57%) and Extended

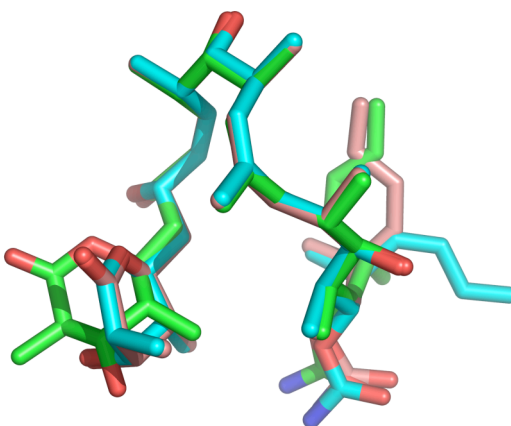
(9%). Seven out of twelve are hydrogen bonded with 5 sustaining a hydrogen bond between a hydroxyl on the main chain or the lactone and the carbamate side chain. Five conformations including NAMFIS-2 (14%) and NAMFIS-3 (12%) display the lactone in a boat conformation with all substituents equatorial. Many of the NAMFIS conformers also share similarity with each other. For instance, NAMFIS-5 (10%), NAMFIS-9 (4%) and NAMFIS-12 (2%) superpose very well on each other (**Figure 2.19**), as do NAMFIS-3 (12%) and NAMFIS-8 (5%). The principal difference between these conformations is found in the butenyl side chain and the orientation of the lactone, whereas for most of them the C5-C18 section overlaps closely.



**Figure 2.19:** Overlap of NAMFIS-5 (blue), NAMFIS-9 (pink), and NAMFIS-12 (green). The average heavy-atom RMSD is 0.5 Å.

Comparison of many of the NAMFIS conformers with the X-ray conformation indicates a surprising similarity between them, especially in the C5-C18 segment. For example the

X-ray conformer superposes very well with NAMFIS-12 (2%), NAMFIS-5 (10%), NAMFIS-4 (12%) as illustrated in **Figure 2.20**.



**Figure 2.20:** Overlap of the X-ray conformation (green) with NAMFIS 12 (pink) and NAMFIS-5 (blue); RMSD = 0.4 Å.

Overall, 8 out of 12 NAMFIS structures overlap with the X-ray conformation in the C5-C18 region with RMSD values of less than 1 Å, and 4 of these overlap with values of less than 0.5 Å. The main conformational variability exists in the butenyl and lactone side chains. As noted above, since the X-ray structure is very similar to the force field global minima, the aforementioned NAMFIS conformers also overlap closely with the global minima. In terms of percentage, 57% of conformers in DMSO- $d_6$  are part of the x-ray family and 42% are part of the non-x-ray family. This dominance of the x-ray family contrasts with the Snyder et al. study in which no more than 1% of the conformations resembled the x-ray conformation. Thus the study indicates that multiple force field

searching mapped the conformational surface for DDM more comprehensively and provided a substantially different result from that from the single force field search. However, similar families are also located in both the studies; the best fit conformation in the present study for instance belongs to the second-most dominant ‘Sickle’ family in the previous study.

The percentage of conformers belonging to the x-ray family increases substantially in water. Using the distance and coupling constant data in water from the Canales et al. paper, we analyzed DDM conformations in water using NAMFIS (SSD 194) and found that 80% of the conformers belong to the x-ray family. However, the number of distinct conformers in water compared to DMSO is also greater; 16 compared to the previous 12. Examination of the conformers in both solvents reveals similarities. For instance an identical conformer from the x-ray family is found in DMSO-d<sub>6</sub> (9%) and D<sub>2</sub>O (2%). Examples of the sickle (17% and 4%) and dome (17% and 13%) families are also found in both solvents. The extended family is present only in DMSO-d<sub>6</sub>. However the x-ray family clearly dominates in both solvents. **Table 2.3** illustrates the percentages of conformers in both datasets and the classification of these conformers in x-ray and non-x-ray families; the criterion used to classify a conformer as belonging to the x-ray family is a C5-C19 heavy atom RMSD of 0.5 Å or less.

Conformer no.	% DMSO-d <sub>6</sub>	% D <sub>2</sub> O	Family in D <sub>2</sub> O	Family in DMSO-d <sub>6</sub>
1	17%	12%	X-ray	Sickle
2	14%	11%	X-ray	Dome
3	12%	10%	X-ray	X-ray
4	12%	10%	Dome	X-ray
5	10%	9%	X-ray	X-ray
6	9%	8%	X-ray	X-ray
7	9%	7%	X-ray	Extended
8	5%	4%	Sickle	X-ray
9	4%	4%	X-ray	X-ray
10	4%	4%	X-ray	X-ray
11	3%	3%	Dome	Dome
12	2%	3%	X-ray	X-ray
13		2%	X-ray	
14		2%	Dome	
15		2%	X-ray	
16		2%	X-ray	

**Table 2.3:** Percentages of DDM populations in DMSO-d<sub>6</sub> and D<sub>2</sub>O and their classification into x-ray and non-x-ray families.

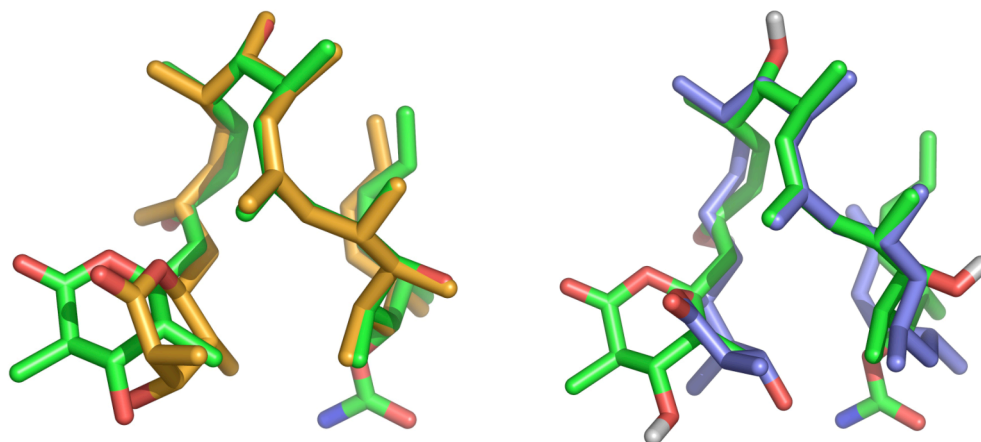


These observations reinforce the observation for DCT; solvent changes, even significant ones from aprotic to protic, result in a redistribution of major conformer families without altering the essential nature of the individual conformations.

As studies in both solvents indicate, clearly there is a dominant and energetically stable conformation for the C5-C18 segment of DDM that emerges in the solid-state, as a global minimum in multiple force field conformational searches and in solution. As discussed below, this dominant motif is also similar to the proposed bioactive conformation.

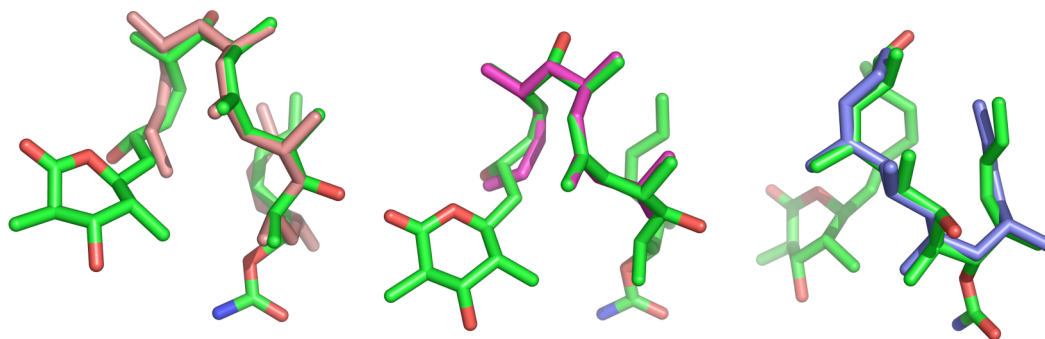
### ***2.10: Simplified DDM constructs***

To try to rationalize the energetic source of the persistent torsional characteristics, we focus on the C5-C18 region and minimize the role of the more flexible carbamate and lactone side chains. Likewise, in order to minimize the role of electrostatic effects that can dominate conformational preferences in force field conformational searches<sup>32</sup>, the carbamate side chain was excised subsequent to evaluating steric aspects of the resulting fragment. In a first calculation we replaced the hydroxyls at positions 7, 11 and 17 with methyl groups. Multiple force-field conformational searches (MM3, MMFFs, OPLS 2005) with this fragment indicated the global minima to be very similar to those of the parent DDM molecule. A similar outcome was achieved by replacing the three hydroxyls with hydrogens. Superpositions of these minima with the X-ray conformation are provided in **Figure 2.21**.



**Figure 2.21:** Overlap of the DDM X-ray conformation (green) with carbamate excised fragment global minima. Left: C-7, C-11 and C-17 hydroxyls replaced by methyl groups; right: C-7, C-11 and C-17 hydroxyls removed.

To further explore the conformational preferences in individual sectors of the molecule, we simplified the conformational profiles and performed OPLS 2005 conformational searches with short non-polar fragments of the discodermolide backbone consisting of the C6-C24, C7-C17 and C12-C24 regions. As before, the global minima for these fragments overlap closely with the corresponding sectors of DDM as shown in **Figure 2.22**.



**Figure 2.22:** Overlaps of the global minima for the C6-C24, C7-C17 and C12-C24 fragments (from left to right) of DDM with the X-ray conformation (green)

The composite results indicate that the energetic preferences for discodermolide arise mainly from steric factors and are preserved in individual segments of the molecule. It is noteworthy that DDM incorporates three methyl-hydroxy-methyl triads which have been exploited as symmetrical synthons in the many convergent syntheses of discodermolide reported<sup>79</sup>. While they are convenient synthetic elements, these synthons likewise seem to be the dominant contributors to the conformation of the molecular backbone. In the further dissection of the energetic characteristics of these synthons, we drew inspiration from the pioneering work of Hoffmann who has investigated *syn*-pentane interactions in substituted pentane and related systems<sup>80,81</sup>. The methyl-hydroxy-methyl fragment can be viewed as a 2,3,4-trisubstituted-pentane system. As such, its conformational preferences are critical for directing the overall shape of DDM and analogs.

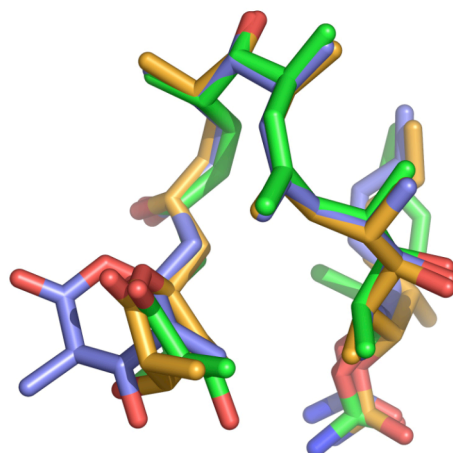
The 2,3,4-trisubstituted-pentane fragment can exist as six potential conformers defined by the relationship of the two methyl groups with respect to the central hydroxyl groups: *a/a*, *a/g+*, *a/g-*, *g-/g-*, *g+/g+* and *g-/g+* (where *g* refers to *gauche* and *a* refers to *anti*). Ab initio quantum chemical single-point energy evaluations of the force-field optimized conformers at the B3LYP/6-31G\* level suggests that *g-/g+* is the most stable conformation for this fragment when the absolute configuration is SSS as for the DDM C10-C12 segment. The energetic difference between this and the next-most stable conformation, the *a/g+*, is 0.1 kcal/mol, with a maximum energy difference of 3.3 kcal/mol between this conformer and the *a/a* conformation. Corresponding calculations for 2,3,4-trisubstituted pentane with an SRS absolute configuration, similar to that for the C16-C18 section of DDM, indicate the dominance of the *g+/g+* conformation with similar energy differences between it and other conformations. While the relatively low value of the individual energy differences may not indicate a well-defined conformational preference, the presence of such centers in the three relevant sections of DDM sums to a pronounced conformational bias in the molecule.

These calculations and hypotheses are supported by an inspection of the NAMFIS conformations. For example for the C10-C12 fragment, 8 of the 12 NAMFIS conformations present the fragment in a *g-/g+* conformation. On the other hand, the best-fit solution conformer (NAMFIS-1) displays this fragment in the *a/g+* conformation which as stated above, is only 0.1 kcal/mol higher in energy than the *g-/g+* conformation. The C16-C18 section of DDM lends itself to similar conformational preferences with 9 of 12 conformers having the C16-C18 sector in the *g+/g+* conformation. Conformational

analysis thus underscores the presence of preferred torsional motifs in the C10-C12 and C16-C18 segments.

### ***2.11: The tubulin-bound conformation of DDM***

Two groups have applied experimental approaches to determine a tubulin-bound conformation for DDM. For example, Sanchez-Pedregal et al. proposed a model for the bound conformation of DDM by means of transfer-NOESY experiments in the presence of epothilone A<sup>76</sup>. The molecular geometry was suggested to be similar to the solid-state and solution conformation of the same ligand. More recently Canales et al.<sup>73</sup> deduced the bound conformation of DDM with similar experiments from which it was concluded that there are only minor differences between this conformation and the originally proposed Sanchez-Pedregal et al. solution conformation, most notably in the disposition of the butenyl group. This study also postulated the existence of a major global minimum conformation in water. Thus, both studies suggest a very similar conformational correspondence between the solid-state, solution and protein bound conformations. By contrast, our investigation indicates a broad distribution of DDM conformations in solution with the proposed bound conformation represented as a dominant geometry among them. **Figure 2.23** exhibits the overlap between the solid-state, NAMFIS-10 and protein-bound conformations (avg. C5-C18 RMSD 0.5 Å).

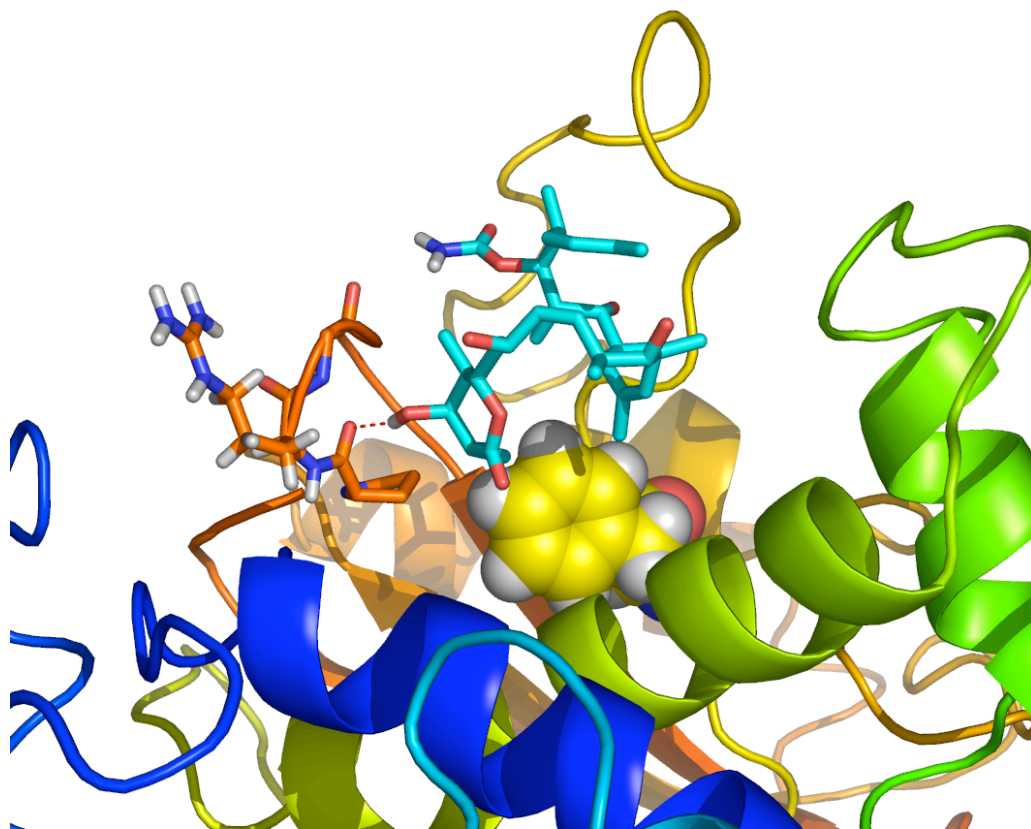


**Figure 2.23:** Overlap of DDM tubulin-bound (golden), X-ray (green) and NAMFIS-10 (blue) conformations.

Interestingly, in the Canales et al. study, the conformer providing the best fit to the NMR variables is a global minimum on the MM3\* potential energy surface. In the present work, the global minima from the AMBER\*, MMFFs and OPLS- 2005 searches were also similar to the MM3\* global minimum. However, in solution, since a non-trivial fraction of the NAMFIS conformations (eg. NAMFIS-1, 2 and 3) are dissimilar from the proposed bound conformation, there would appear to be a much greater degree of conformational selection during the binding event than suggested by the previous two reports.

The Canales et al. paper also included docking studies with DDM and based on some SAR data, proposed a binding pose of DDM in the tubulin binding pocket. The program AUTODOCK was used to generate the binding pose. We decided to repeat the docking

experiments both to check the validity of the binding pose and to investigate the possible existence of better poses. To this end GLIDE (v 3.5) was used to generate 20 poses by docking DDM rigidly in the surmised conformation into the tubulin (PDB code: 1JFF) binding site. All settings used were default, including settings for protein preparation, generation of a grid centered on the ligand and docking in Standard Precision (SP) mode. SAR data for DDM has often emerged as contradictory and weak, but one consistent conclusion from many SAR studies is the non-involvement of two residues, Phe 270 and Ala 234 in the binding of DDM; this fact is suggested by the conservation of full DDM activity in tubulin harboring the F270A and A234T mutations which abolish activity of taxol<sup>76</sup>. A pose very similar to the previous AUTODOCK pose emerges as the 13<sup>th</sup> best pose ranked by GLIDE. To gain a more realistic appraisal of the poses, we decided to use the MM-GBSA protocol to re-rank the poses since MM-GBSA takes account of solvation of the protein ligand complex and the individual protein and ligand using an implicit solvation model (GBSA) and aims to more accurately represent the physical chemistry of ligand binding. It has been used recently to re-rank poses from virtual screening<sup>40</sup>. Surprisingly, MM-GBSA scored the previous 13<sup>th</sup> ranked GLIDE pose as the top pose, with an estimate relative  $\Delta G$  of 2 kcal/mol between it and the next-best pose. This ranking further augmented our confidence in the orientation of the proposed binding mode. The MM-GBSA pose is displayed in **Figure 2.24**.



**Figure 2.24:** The best binding pose of DDM in tubulin as obtained by GLIDE docking followed by MM-GBSA re-docking. The lactone C3-hydroxyl forms a hydrogen bond with a proline carbonyl. Phe 270 rendered in space-filled form does not have measurable contacts with DDM.



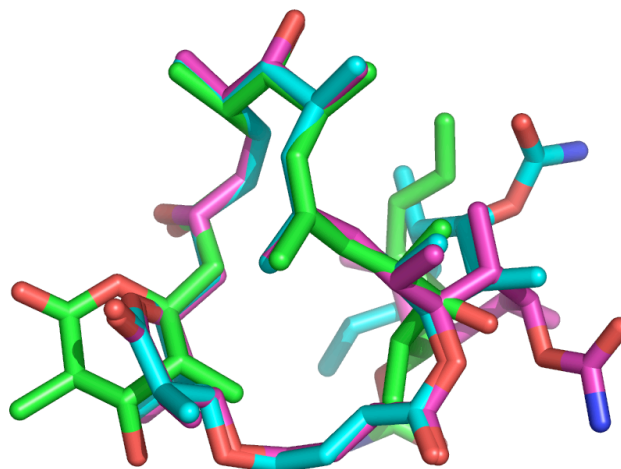
### ***2.12: Conclusions and outlook***

In summary, a multiple force field and NAMFIS solution analysis of discodermolide has been performed to provide a realistic and high quality fit of the NMR data obtained in DMSO to an ambient solution ensemble of conformations. The molecule adopts a preferred conformation, consistent with the conformation advocated for both the X-ray crystal structure and the proposed tubulin-bound form. Conformational analysis aided by force field and quantum chemical calculations underscore the notion that steric factors stabilize the dominant conformation of DDM in three separate environments: solution, the solid-state and in the tubulin binding site. Studies on flexible, bioactive molecules from our laboratory have suggested that the ligand X-ray structure is commonly influenced by crystal packing factors. As a result, it is often a minor component, if present at all, of the conformational ensemble in solution where such packing effects are absent<sup>36</sup>. In general, NAMFIS solution studies on other bioactive ligands binding to tubulin such as taxol<sup>17</sup> and the epothilones<sup>33</sup> have identified the bioactive conformation as a minor solution ensemble component.

Discodermolide would appear to be a significant exception to the latter generalization due to the presence of the three rigidifying fragments containing tandem vicinal Me, OH and Me functionality. Relatively straightforward steric factors seem to influence the existence of a common dominant conformer in solution, the solid state and in the protein. This is rarely the case, since differential energetic requirements ordinarily lead to conformational diversity among the micro-environments<sup>36</sup>. In the present case however, the overriding conformational preference is induced largely by steric factors that lead to a

dominant 3D form under all three circumstances. Insofar as the conformational preferences of DDM dictate the latter under very different physical conditions, the methyl-hydroxy-methyl triad and related moieties suggest themselves as modular, lego-like elements that permit straightforward application of molecular mechanics-guided conformational principles to prediction of the bioactive conformations of related natural and synthetic products. Such approaches may also suggest synthetic design strategies incorporating multiple appearances of the modular component to impose conformational preferences.

Based on hypotheses about the dominant conformation, we are collaborating with David Kingston's group at Virginia Tech<sup>82</sup> to examine the feasibility of modified DDM analogs that are constrained into shapes resembling the dominant conformation. To this end we have performed conformational searches on the constrained analog illustrated below (**Figure 2.25**) using multiple force fields. These searches indicate global minima from multiple force fields to overlap well with the above global minima of DDM from corresponding searches as well as the bound DDM conformation. Based on the above analyses regarding global minima for DDM being similar to the DDM bound conformation, we conjecture that the constrained DDM analog should preserve the shape of DDM in the tubulin binding site but should potentially show enhanced activity since the constraints would ensure that it would not have to pay the entropic penalty that DDM has to pay in order to bind. Studies concerning the synthesis and biological testing of this analog are underway in the Kingston lab and in Susan Bane's lab at SUNY-Binghamton.



**Figure 2.25:** Constrained DDM analog MMFFs global minimum (cyan) superposed with DDM bound conformation (green) and DDM MMFF global minimum (pink). RMSD is 0.5 Å

The combined conformational and NAMFIS analyses also indicate that simplified versions of discodermolide, for example ones in which the hydroxyls are removed or replaced by methyls or ones in which certain portions of the chain are excised or substituted by other groupings, may preserve the DDM conformational preferences. Such simplified constructs, by preserving binding interactions driven by the adoption of a favorable molecular shape, can possibly provide improved drug-like properties, thereby enhancing therapeutic utility.

### ***2.12: Perspective: Conformations of tubulin-binding agents***

Tubulin-binding ligands represent some of the most diverse chemistry found within ligands binding to a single protein. The chemistry of these ligands is also reflected in their biological action, with some of them acting as microtubule-stabilizing agents and others acting as destabilizing agents and with many of them producing divergent morphological and kinetic changes in tubulin assembly and disassembly. The diversity in chemistry is also reflected in their binding patterns with laulimalide and peloruside thought to bind to a site different from that occupied by DDM, DCT, taxol and epothilone. During the past decade our group has analysed the solution, solid-state, and wherever possible, binding conformations for many of these molecules. In case of taxol and epothilone the bioactive conformation has been located as a minor constituent of the solution ensemble.

The present studies on DCT and DDM illustrate both similarities and differences with these previous studies. DCT seems to mirror taxol and epothilone in presenting the bioactive conformer as a minor conformation (7%) in solution. At the same time this conformer, while belonging to the same family as the proposed bioactive conformer, also shows distinctive differences from the latter, and may possibly provide a better fit to the NMR data obtained from tr-NOESY experiments. As such there seems to be a diverse conformational ensemble that DCT adopts in various solvents.

The situation with DDM seems to be drastically different. In this case there seems to be a dominant conformation for the molecule that's remarkably prevalent in the solid-state, in

multiple force field conformational searches, in three different solvents and in the protein binding site. This conformation seems to be reinforced by pronounced steric influences that seem to override other preferences influenced by electrostatic, polar and solvent interactions. Yet at least one variable, the role of solvent, seems to be clear with water influencing an observable change in the dominant conformation (80%) compared to DMSO (58%). Nonetheless, the almost ubiquitous conformational preferences exhibited by this molecule in diverse microenvironments seems to bear out an essential component of synthetic molecular design; that of rationally designing molecules with selective shapes by carefully modulating the steric and polar interactions in their building blocks. Thus, while the divergent behavior of DCT and DDM demonstrate the remarkably broad chemistry exhibited by tubulin-binding agents, the analyses of DDM conformations also provides a direction in designing custom-made molecular fragments that could regulate the activity of this important and endlessly interesting protein in health and disease. As hinted by the constrained analog of DDM, such promising investigations are underway.

Finally, the rather dramatic differences between the conformational behavior of DCT and DDM again reveal the complexities of 3D structure that 'similar' 2D structural comparisons may mask, and validate the value of NAMFIS-like methods for exploring these complexities.

# Chapter 3

## Noscapine

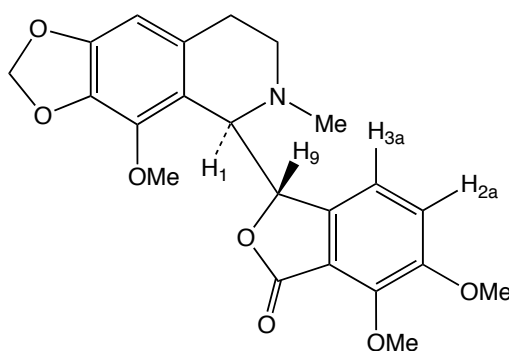
“No problem is too small or too trivial if we can really do something about it”

-Richard Feynman

### 3.1: Introduction

The opium poppy, a flowering plant that has been used for centuries for recreation, culinary and decorative purposes, is best known for containing morphine<sup>83</sup>.

Morphine constitutes about 80-90% of the opium poppy seed. Noscapine (**Figure 3.1**), an alkaloid that is a minor constituent of the poppy seed (1-10%), is available over the counter as an antitussive agent<sup>84</sup>.



**Figure 3.1:** Noscapine

It has been used for many years in this capacity and has benefited from a lack of significant side effects even in large doses. Recently, noscapine and its analogs have been shown to interact with microtubules and to destabilize them in a manner reminiscent of various microtubule destabilizing agents<sup>85,86</sup>. Noscapine shows good oral bioavailability<sup>87</sup> and has shown to be effective against H460 NSCLC cells in nude mice. An analog of noscapine, EM105, is more potent than the parent compound and regresses breast cancer xenografts in nude mice<sup>88</sup>. Due to these properties, noscapine and its analogs are under study as anticancer agents and at least one company, Cougar Biotechnology, is studying

these compounds in Phase I clinical trials<sup>89</sup>. It is hoped that the known low toxicity of noscapine will render these leads as viable anticancer drugs.

However, the binding site and binding conformation of noscapine are not known. Earlier it was thought that noscapine and colchicine may share a similar binding site on the protein because of similarities in chemical structure, but labeling and competition studies indicated otherwise<sup>90</sup>. An investigation of the low-energy conformations that the compound adopts in solution may prove useful for such a study as it has for paclitaxel<sup>91</sup> and epothilone-A<sup>33</sup>. To this end, a conformational analysis of **1** and its hydrochloride salt (1-HCl) in solution was performed by using a combination of molecular mechanics-based conformational searching, and measurements from one and two-dimensional NMR spectroscopy. NAMFIS was employed to deconvolute the average NMR data into a population ranking of low-energy conformers in solution. A prominent feature emerging from this work is the dominance of a pseudoaxial N-methyl group in a number of the conformers in solution and in the X-ray structure of the noscapine salt<sup>92</sup>. Application of force field and quantum chemical methods provides a rationalization of this conformational preference.

### ***3.2: NOESY Spectra of Noscapine***

NOESY spectra of noscapine (**1**) and its N-protonated hydrochloride salt (1-HCl) were acquired on a Varian INOVA 600 MHz spectrometer at 298 K. The sample was prepared by dissolving 17 mg of noscapine or the salt in 0.7 ml of CDCl<sub>3</sub> and DMSO-*d*<sub>6</sub>. NOESY spectra were recorded at 300, 400, 500, and 600 ms mixing times to check the linearity of



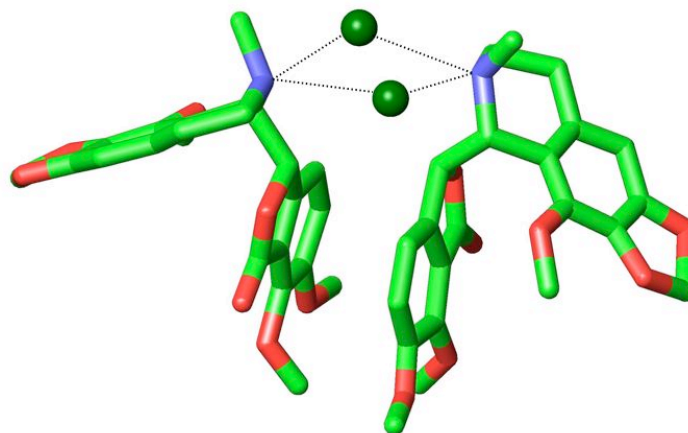
the cross-relaxation build-up rate. Interproton distances were calculated at 500 ms for the neutral compound and at 300 ms for the salt, using an internal calibration distance of 1.8 Å between the geminal protons adjacent to nitrogen. Relaxation delay was set to 1.5 s for noscapine and to 8 s for the salt. 10 NOE derived distances for the neutral compound, 20 NOE distances for the hydrochloride, and one critical inter-ring  $^3J_{\text{H-H}}$  spin-spin coupling constant between H-1 and H-9 were obtained.

To generate a conformation pool of structures, conformational searching for 1 was carried out with three force fields implemented in the Macromodel program embedded in the Maestro suite from Schrödinger Inc.: MMFF, AMBER\* and MM3\*. The GBSA/H<sub>2</sub>O and GBSA/CHCl<sub>3</sub> continuum solvent models were used for all conformational evaluations. Each search was performed with the Mixed Torsional Search/Low-Mode Monte Carlo Minimum method with a 7 kcal/mol cutoff for 5000 steps. Since Noscapine contains only 6 rotatable bonds, the latter ensured that the entire potential energy surface of the molecule was explored<sup>26</sup>. In confirmation, every search located the global energy minimum at least a hundred times. The 7 kcal/mol cutoff was introduced for all conformational searches to avoid generation of high-energy conformers. Finally, the structures from each conformational search were optimized with the force fields used for the respective searches to provide full energy convergence. The search results from the three force fields were pooled together, and duplicates were eliminated. These manipulations yielded 17 and 22 unique conformers for the neutral compound and hydrochloride salt, respectively. The final ensembles of structures were used as inputs to the NAMFIS analysis.

Post-NAMFIS evaluation assists in eliminating conformations that are chemically unreasonable, but nonetheless fall within the 7 kcal/mol window employed during the conformational searches. In the present case, the conformers obtained from the NAMFIS analysis were subjected to single point energy calculations with the B3LYP/6-31G\*\* DFT model using the Jaguar program from Schrödinger<sup>93</sup>. For comparison, DFT single point calculations were also performed on X-ray crystal structures of 1 and 1-HCl (see below). Before performing the DFT calculations, both NAMFIS conformers and crystal structures were optimized using the MMFF force field and frozen torsional angles throughout to ensure that all conformers lie on the same potential energy surface.

### ***3.3: Noscapine X-ray crystal structures***

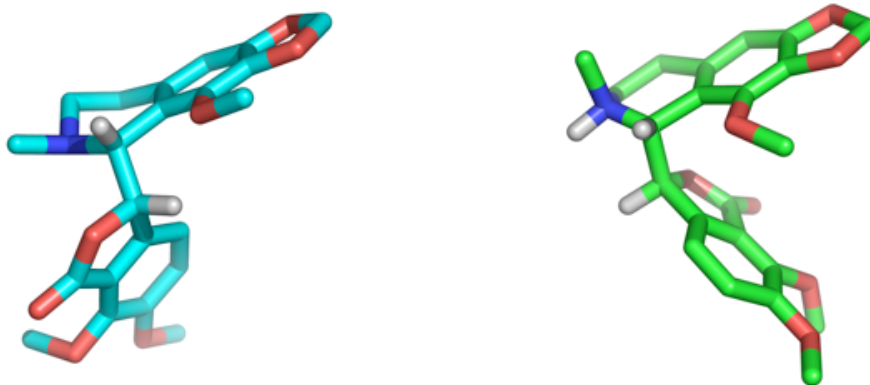
A crystal structure of the HCl salt of noscapine has been reported<sup>94</sup>. We have also independently obtained the crystal structure of neutral noscapine. In the solid state, noscapine hydrochloride (1-HCl) contains two molecules in the unit cell. The N-Me groups are pseudoaxial, and two chloride ions form pseudoequatorial Cl---NH bridging interactions between the two molecules (**Figure 3.2**).



**Figure 3.2:** Two molecules of noscapine hydrochloride (1-HCl) in the crystal unit cell. The green spheres are chloride ions which form a bridging interaction between the two pseudo-equatorial N-H moieties.

### ***3.4: Conformer classification***

It is convenient to classify noscapine conformers on the basis of two parameters: pseudoaxial or pseudo-equatorial disposition of the N-Me group, and the value of the H1-C1-C9-H9 dihedral angle (See structure 1). In principle, there are three distinct conformations about the latter bond; two gauche and one anti. The X-ray structure of 1-HCl displays a value of  $+75^\circ$  for this dihedral angle, and the N-Me group is pseudoaxial. The structure of neutral noscapine, on the other hand, sustains H1-C1-C9-H9 at a value of  $-66^\circ$  and presents a pseudo-equatorial N-Me group. The two crystal structures are compared in **Figure 3.3** with the benzopyridine rings in the same orientation.



**Figure 3.3:** The solid state conformations of noscapine. Left: neutral **1** (blue),  $\angle$  (H1-C1-C9-H9) =  $-66^\circ$ ; Right: **1-HCl** (green),  $\angle$  (H1-C1-C9-H9) =  $+78^\circ$ .

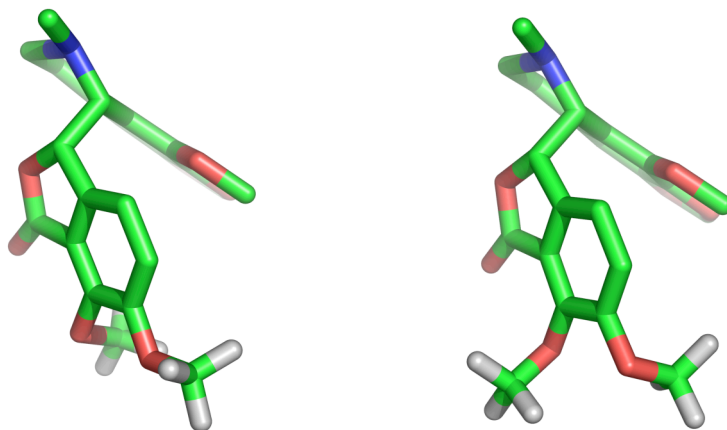
### ***3.5: Noscapine hydrochloride conformations in solution***

Triple force field analysis for **1-HCl**, excluding the chloride ion, delivered 22 unique optimized conformers. These structures along with 20 NOE-determined distances and one cross-ring  $^3J_{\text{H-H}}$  coupling constant were integrated by the NAMFIS protocol to provide five conformers representing the solution average. Estimated populations are given in **Table 3.1**. The free energies derived from Boltzmann treatment of the populations fall within 2.2 kcal/mol of the global minimum (**1-HCl-1**, population 41%).

Conformer	NAMFIS population, %	$\Delta G_{\text{rel}}$ , kcal/mol (298 K)	N-Me, pseudo-	$\angle$ (H1-C1-C9-H9) deg
1	41	0.0	ax	78
2	27	0.3	ax	-40
3	16	0.6	eq	74
4	14	0.6	ax	78
5	1	2.2	eq	74

**Table 3.1:** Geometries and Boltzmann energies of the NAMFIS conformers of nescapine hydrochloride salt 1-HCl in  $\text{CDCl}_3$ .

NAMFIS analysis posits the X-ray conformation of 1-HCl to be present in solution as the top conformer (1-HCl-1, 41%) with a H1-C1-C9-H9 dihedral angle of  $+78^\circ$  and a pseudoaxial N-Me group. The second most populated conformer (1-HCl-2, 27%), like the first and fourth (1-HCl-4, 14%), presents its N-methyl in the pseudoaxial orientation. Thus the pseudoaxial conformer dominates the ensemble with a total population of 82%. The third and fifth best fits (1-HCl-3, 16% and 1-HCl-5, 1%, respectively) direct N-Me to be pseudoequatorial and share a common cross-ring torsion angle ( $+74^\circ$ ). The single difference between them is found in the disposition of the two adjacent methoxy groups on the lower benzene ring in **Figure 3.4**. A similar situation obtains for 1-HCl-1 and 1-HCl-4.



**Figure 3.4:** The third and fifth NAMFIS conformers (1-HCl-3 and 1-HCl-5) differ only in the orientation of the two OMe groups on the lower benzene ring.

Relative energy calculations were performed for the five positively charged conformers with both molecular mechanics (MMFF) and DFT (B3LYP/6-31G\*\*, single point for MMFF optimized) (**Table 3.2**). The GBSA/CHCl<sub>3</sub> solvation model was employed for the MMFF calculations while DFT values were obtained using the PBSA/CHCl<sub>3</sub> solvation treatment implemented in the Jaguar program. Both methods correctly identify NAMFIS 1-HCl-1 as the lowest energy, dominant conformer. However, the same energy evaluations do not predict the empirical ranking of the NAMFIS conformers, while the DFT calculations seriously overestimate energies of the low population isomers by comparison with the NMR/NAMFIS result. A similar scenario has been observed for conformations of laulimalide<sup>36</sup>. On the other hand, the MMFF force field energies all fall within 2.0 kcal/mol of the lowest energy conformer in accord with the NAMFIS relative energies.

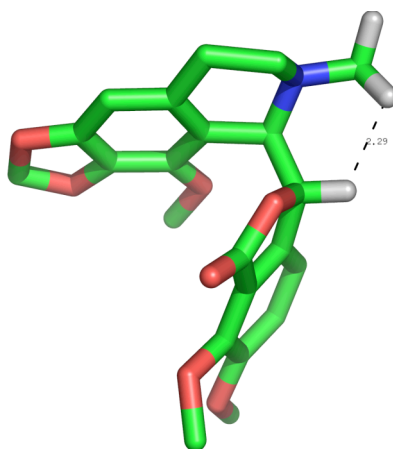
NAMFIS	$\Delta G_{\text{rel}}$ , (300 K)	MMFF/GBSA/ CHCl <sub>3</sub> , $\Delta E_{\text{rel}}$	B3YLP/6-31G**/PBSA/ CHCl <sub>3</sub> , $\Delta E_{\text{rel}}$
population, %			
41	0.0	0.0	0.0
27	0.3	0.9	4.4
16	0.6	1.7	3.9
14	0.6	1.4	3.0
1	2.2	0.3	0.1

**Table 3.2:** Relative energies of 1-HCl NAMFIS conformers, kcal/mol.

In an attempt to explain why the dominant 1-HCl conformer orients its N-Me in a pseudoaxial position, 1-HCl-1 and the identical structure with the N-Me group oriented in a pseudoequatorial orientation were optimized with the MMFF and OPLS-2005 force fields. Torsional constraints were used for the equatorial counterpart so that only the axial-equatorial conformational variable would be probed. This procedure leads to the result that pseudoequatorial 1-HCl-1 is estimated to be about 0.5 kcal/mol higher than pseudoaxial 1-HCl-1. The pseudoequatorial form is in fact the third best fit conformer 1-HCl-3. Not surprisingly, the difference in  $\angle$  (H1-C1-C9-H9) between the two is only 4° (78° vs 74°). Thus, the equatorial counterpart of the best fit axial conformer is identified by NAMFIS as nearly isoenergetic.

As model systems for evaluating the similar energies of 1-HCl-1 and 1-HCl-3, we have examined MMFF-optimized 3-methyl cyclohexene and N-protonated N-methyl tetrahydropyridine, structures lacking a large substituent such as the annulated lactone in

1 and 1-HCl. The MMFF/GBSA/CHCl<sub>3</sub> energy differences between the pseudoaxial and pseudoequatorial conformers are 1.2 and 1.4 kcal/mol respectively favoring the equatorial forms. A possible implication is that steric congestion present in 1-HCl, but absent in the model structures is responsible for the energy compression between 1-HCl-1 and 1-HCl-3 (**Figure 3.5**)



**Figure 3.5:** Steric clash in the 1-HCl-3 conformer of protonated noscapine, the pseudoequatorial counterpart of the NAMFIS global minimum 1-HCl-1 (41%). The MMFF N-CH<sub>3</sub>H---H<sub>9</sub> interproton distance is 2.3 Å.

Indeed, 1-HCl-3 incorporates a short contact distance (2.3 Å) between an N-Me proton and the lactone ring H<sub>9</sub> proton (**Figure 3.5**) that is just under the sum of the Van der Waals radii.<sup>1</sup> Such an interaction could raise the energy of this conformation to within 0.5 kcal/mol of 1-HCl-1. This cannot be the whole story, however, since the latter sustains a 2.4 Å distance between the NH proton and H<sub>9</sub> in the lactone. Thus, while we cannot definitively account for the diminutive energy difference between 1-HCl-1 and 1-HCl-3,



we surmise that steric effects regulated by small differences in intramolecular contacts are the basis for raising the energy of the pseudoequatorial conformer relative to the pseudoaxial isomer. A similar analysis for the second pseudoequatorial conformer (1-HCl-5, 1%) suggests that an analogous steric interaction is most likely responsible for its estimated low abundance in solution compared to the pseudoaxial conformers.

### ***3.6: Neutral noscapine conformers in solution.***

Similar to the procedure for the noscapine salt 1-HCl, a three-way force field conformational analysis (MMFF, MM3 and AMBER\*) yielded 17 unique conformers for the neutral counterpart. Subsequent NAMFIS deconvolution of the NMR spectra of noscapine in DMSO and chloroform delivered three and four conformers respectively. Free energies of the solution forms calculated from their Boltzmann populations are within 2 kcal/mol of the corresponding global minima. Solution populations and selected structural features of the equilibrating forms in the two solvents are listed in **Table 3.3**.

Conformers	NAMFIS population, %	$\Delta G_{\text{rel}}$ , kcal/mol (300 K)	N-Me. pseudo-	$\angle$ (H1-C1-C9-H9), deg
DMSO- <i>d</i> <sub>6</sub>				
1	71	0.0	ax	-48
2	27	0.6	eq	-55
3	6	1.5	eq	71
CDCl <sub>3</sub>				
1	54	0.0	ax	-48
3	26	0.4	eq	71
4	13	0.8	eq	77
5	7	1.2	ax	178

**Table 3.3:** Geometries and Boltzmann energies of the NAMFIS conformers of noscapine (1) in DMSO-*d*<sub>6</sub> and CDCl<sub>3</sub>.

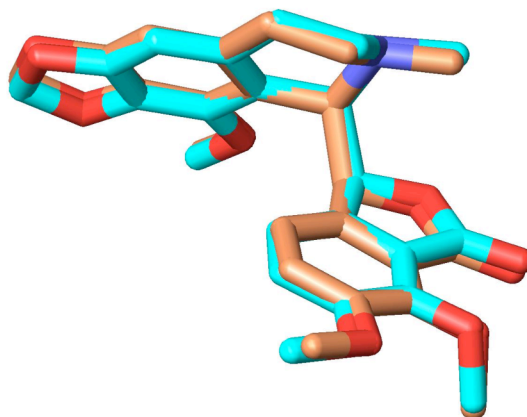
Analogous to 1-HCl, a pseudoaxial conformer with  $\angle$  (H1-C1-C9-H9) = -48° (ax-1-1) is dominant in both solvents. To probe this result, the corresponding equatorial conformer was MMFF-optimized with torsional constraints identical to those for ax-1-1. The difference in energy between this virtual pseudoequatorial conformer and 1-1 is 2.6 kcal/mol, while  $\angle$  (H1-C1-C9-H9) settled at -50°. The distance between one of the N-Me protons and the lactone ring oxygen is 2.51 Å, 0.2 Å short of the H---O van der Waals sum<sup>95</sup>. The relatively short distance introduces a steric clash which most likely leads to the higher energy of the virtual pseudoequatorial rotamer. Nonetheless, this conformer

shares properties with the most populated equatorial conformer in DMSO (eq-1-2, 27%) with  $\angle$  (H1-C1-C9-H9) =  $-55^\circ$ . But after MMFF optimization, the  $\angle$  (H1-C1-C9-H9) angle in eq-1-2 (the most populated equatorial conformer) changes to  $-62^\circ$ . In this conformation, the N-Me-lactone ring oxygen distance is now 2.62 Å, slightly longer than in the virtual equatorial form. Once again, a small difference in the Van der Waals sum may be responsible for the higher calculated energy difference (2.6 kcal/mol) between the virtual conformer and pseudoaxial 1-1, compared to the lower energy difference (0.6 kcal/mol) between eq-1-2 with an expanded torsional angle ( $-55^\circ$ ) and ax-1-1 ( $-48^\circ$ ).

In chloroform, while ax-1-1 is the global minimum, eq-1-2 does not appear following NAMFIS treatment. Instead, the next most populated conformation, pseudoequatorial 1-3, is very similar to the most populated equatorial structure indicated in the same solvent for 1-HCl (i.e. eq-1-3, **Table 3.1**). Similar to the monocyclic model salt discussed above, the MMFF/GBSA/CHCl<sub>3</sub> energy difference between the axial and equatorial conformations of neutral N-methyl tetrahydropyridine is 2 kcal/mol. Neutral noscapine (1), like its salt 1-HCl, seems to suffer steric crowding with the appended lactone ring in the pseudoequatorial geometry, therefore raising its energy and permitting the pseudoaxial conformation to effectively compete as an equilibrium partner.

It is noteworthy that the most populated equatorial conformer in DMSO is very similar to that captured in the solid state (Figure 2). In fact, optimization by MMFF of both this conformer and the crystal conformer changes  $\angle$  (H1-C1-C9-H9) to  $-62^\circ$  and drops both conformations into the same energy minimum. Even before optimization, the two

conformers with dihedral angles  $-55^\circ$  and  $-66^\circ$  are essentially identical, as shown in the following superposition (**Figure 3.6**)



**Figure 3.6:** Superposition of the x-ray conformation of neutral noscapine (blue carbons) and the most populated equatorial conformer in DMSO (brown carbons)

An energetic evaluation for the conformers was performed using both molecular mechanics and DFT (B3LYP/6-31G\*\*) methodologies using the GBSA/CHCl<sub>3</sub> and PBSA/CHCl<sub>3</sub> models, respectively. (**Table 3.4**)

NAMFIS population, %	$\Delta G_{\text{rel}}$ (300 K)	$\Delta E_{\text{rel}}$ , MMFF/ GBSA/CHCl <sub>3</sub>	$\Delta E_{\text{rel}}$ , B3YLP/ 6-31G**/PBSA
DMSO- <i>d</i> 6			
71	0.0	0.0	0.0
27	0.6	2.3	2.1
6	1.5	-0.5	1.7
CDCl <sub>3</sub>			
54	0.0	0.0	0.0
26	0.4	0.5	2.6
13	0.8	0.5	2.4
7	1.2	1.2	2.4

**Table 3.4:** Relative energies of noscapine (1) NAMFIS conformers, kcal/mol.

For chloroform, the agreement between NAMFIS and MMFF energies is impressive and superior to the same analysis for 1-HCl (**Table 3.2**). In the case of DMSO, the force field cannot correctly rank the NAMFIS global minimum. However, the error, 0.5 kcal/mol, is within the limits of the accuracy of current force fields. DFT calculations identify the lowest energy NAMFIS conformers in both solvents, but somewhat overestimate the NAMFIS relative energies. The failure to correctly rank the conformers is within 0.4 kcal/mol, however, an energy window that DFT cannot accurately parse<sup>96</sup>.

From the above analysis, it can be concluded that the same factors controlling stability of the pseudoaxial N-Me orientation in the noscapine salt (1-HCl) favor the pseudoaxial over the pseudoequatorial conformer in neutral 1; namely steric interactions between the N-Me center and the lactone ring. To further validate this general conclusion, the lactone ring was excised from both neutral 1 and protonated 1-HCl. DFT (B3LYP/6-31G\*/PCM, Water) single point energy evaluations were subsequently performed for the corresponding MMFF optimized tetrahydro isoquinoline structures. Within this context, the equatorial conformer was indeed more stable than the axial form by 1.7 kcal/mol and 1.5 kcal/mol, respectively. We take this to confirm that torsional variations of the bond connecting the two ring systems cannot eliminate a steric clash between them, and that these steric clashes lead to relatively higher energies for equatorial conformers in both compounds. It is important to note, however, that the calculated  $\Delta E$  (ax-eq) values between the dominant isomers are small and within the error limits of current force fields. More reliable are the NAMFIS free energy differences, but even here a  $\Delta G = 0.5$  kcal/mol can be interpreted as close to the boundary of error limits.

As to preferred geometry, it is clear that the conformational preference of the N-Me group depends on the dihedral angle between the two rigid noscapine ring systems. For most of the conformers,  $\theta$  (H1-C1-C9-H9) is about 75 degrees. At this value, the axial conformer appears to be preferred by a small energy difference.

### ***3.7: Comparison of solution and solid-state conformations***

The solid-state conformation of protonated 1-HCl presents as a dimer with pseudoaxial N-Me groups and chloride anion bridges between pseudoequatorial N-H centers (**Figure 3.1**). Given that axial methyl groups are infrequently observed in saturated and mono-unsaturated six-membered rings<sup>21</sup>, from the X-ray structure alone it might be concluded that the N-Me orientation in the salt arises from a conformational bias due to the double N-H---Cl hydrogen bonds. While the latter undoubtedly involve a positive stabilizing effect in the crystal, the conformational conclusion would be inaccurate in view of the current NMR solution studies. In CDCl<sub>3</sub> solution, NAMFIS studies unambiguously identify the axial conformation as the most stable form. This observation carries over to the neutral form in both CDCl<sub>3</sub> and DMSO. The origin of the pseudoaxial methyl group in noscapine is clearly a consequence of the inability of the two rigid ring systems tethered by a single bond to avoid steric compression. This effect is obtained in the solid state as well. Consequently, we interpret the pseudoequatorial chlorides in the crystal as orientation by default. Namely, the conformation of 1-HCl captured in the crystal lattice is directed by steric effects to position N-Me axial. As a result, the chloride anions associate comfortably with the pseudoequatorial NH groups, but do not actively cause the axial placement of the N-Me groups. In the case of the neutral version of noscapine, the crystal conformation with the N-Me pseudoequatorial is found in the NAMFIS list, even if not as the best fit conformer.

However, in general, it is difficult to quantify all the packing factors influencing particular conformations in crystals. Even though the rationalizations for the

conformational preferences of noscapine in solution may apply to the crystal conformers, additional packing forces could tip the balance in favor of the axial conformations<sup>97</sup>.

### ***3.8: Summary and outlook***

A multi-conformational force field and NMR analysis of the neutral and HCl salt forms of noscapine in solution reveal the existence of low-energy conformations with the N-methyl group pseudoaxial. This result is complemented by a pseudoaxial disposition of the N-Me group in the solid state as well. Energy calculations indicate that steric factors are most likely responsible for influencing the conformations of the N-Me group in both solution and the solid state. In the solid state, additional packing factors may also influence the conformation preferences of the molecules.

If one of the low energy conformations of noscapine is the binding conformation in its interaction with tubulin, then this axial N-methyl may possibly contribute to the potency of the binding. For example, a bioactive conformation in which the NH hydrogen bonds to an amino acid residue could have the N-Me axial. Such a conformational disposition could lead to favorable interactions with the binding pocket. Future studies that seek to identify the bioactive conformation of noscapine in the tubulin binding site will certainly help to elucidate this phenomenon.



# **Chapter 4**

## **The Stevastelins**

“To see what is in front of one’s nose needs a constant struggle”

-George Orwell

#### 4.1: Cyclic peptides:

Peptides are among the most important molecules involved in the maintenance of function in living systems<sup>98</sup>. They are widely involved in signaling, metabolic regulation and the immune response. Some such as glucagon, insulin, enkephalin, somatotrophin and oxytocin are secreted in the bloodstream at sub-micromolar concentrations and yet exert profound physiological effects on events like blood glucose regulation, reproduction sleep and hormonal regulation<sup>98</sup>. Other peptides like amyloid peptide A $\beta$  are detrimental to the normal workings of the brain when present as aggregates and are key components of plaques in Alzheimer's disease<sup>99</sup>.

Because of their key and potent physiological actions, peptides have received substantial attention as potential drugs. Well-developed and versatile chemical methods for synthesizing natural and modified peptides make such an endeavor additionally attractive<sup>100</sup>. Some naturally occurring peptides are already employed in a therapeutic capacity. Anti-microbial peptides are of special interest in this respect<sup>101</sup> because of the role they play in defending microbes against other pathogens and a few anti-microbial peptides have now been adopted for protecting the human body against infections. For example the antibiotic gramicidin is a peptide isolated from the bacterium *Bacillus brevis* and is widely used in topical applications against Gram-positive and select Gram-negative bacteria.

Unfortunately the use of peptides as drugs is thwarted by their poor bioavailability and instability<sup>102</sup>. Oral administration of peptides suffers from rapid elimination by the acidic

environment of the stomach and more importantly by abundant extracellular non-specific and specific proteases in the bloodstream. To circumvent these problems several strategies are being pursued<sup>103</sup>. Peptide pro-drugs delay the recognition and destruction of the peptides by modulating their chemical structure. The use of non-natural and d-amino acids in peptides has also been used to prevent recognition by proteases. In recent years,  $\beta$ -peptides have generated intense interest as modified peptides exhibiting novel folds and protease-resistant biological activity<sup>104,105</sup>. New delivery systems are being worked out for targeted delivery of peptides to the relevant organs. Finally, peptidomimetics are small molecules that mimic peptide structure and function; these are being extensively explored and developed. The structure-based design of Aliskiren, a renin inhibitor, is a particularly successful example of such development<sup>106</sup>.

One important strategy for modifying peptides and improving their activity concerns the use of cyclic peptides. Cyclization can render bioactive peptides extremely resistant to digestion, allowing them to survive intact in the digestive tract. A variety of naturally occurring cyclic peptides are now known. For instance, daptomycin is a natural lipopeptide synthesized by a *Streptomyces* strain that is effective against several Gram-positive bacteria<sup>107</sup>. One of the most well known cyclic peptides is Cyclosporin, isolated from the fungus *Tolypocladium inflatum* Gams. Cyclosporin binds to the protein cyclophilin, a key player in the immune response<sup>108</sup>. It has been widely used as an immunosuppressant peptide administered to organ transplant patients throughout the world for suppressing the rejection of the donor organ by the immune system<sup>109</sup>.

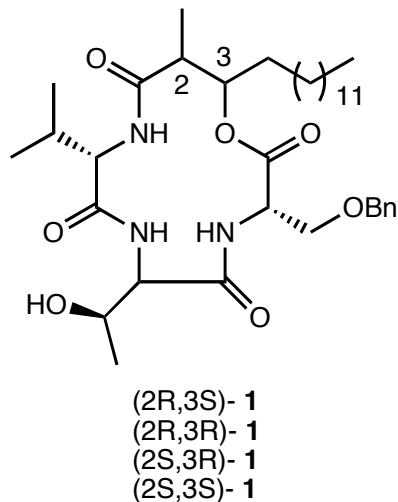
From a design standpoint, cyclic peptides are attractive targets because they are conformationally constrained. The rigidification engendered by these constraints can not only make them serve as mimics of constrained turns in protein structures but also provides the opportunity to fine-tune their conformational, chemical and biological characteristics by making relatively minor, well-defined changes in sequence<sup>110</sup>. Such fine-tuning has been used for influencing substantial changes in biological activity and for increasing oral bioavailability. Knowledge of the constrained conformation can aid in the design of well-defined conformational mimics. From a structural standpoint, small cyclic peptides with four or five residues are hypothesized to be rigid analogues of their linear counterparts because of the ring strain introduced by cyclization. In this respect, 13-membered rings are more interesting since 12-membered rings are believed to suffer from excessive ring strain that renders them relatively unstable<sup>111</sup>.

Even with constraints, cyclic penta- and tetra- peptides still contain several rotatable bonds, and it would be an instructive exercise to investigate the extent of rigidification imposed by the constraints. NMR spectroscopy has been the method of choice for many such past investigations<sup>112</sup>. In typical studies, temperature and solvent-dependent changes in N-H proton chemical shifts are studied to gauge the involvement of these groups in hydrogen bonding and to thence shed light on conformation. Observation of signals for hydrogen bonds corresponding to two different motifs can sometimes provide evidence of interconverting conformations. Yet in the absence of explicit deconvolution techniques, such NMR data is necessarily an average and many previous studies have

relied on simulated annealing wherein a single structure or structural family is varied to fit the average NMR data<sup>112</sup>.

Computational investigations have helped to shed valuable light on such studies. A previous study from our group used NAMFIS analysis to explore the conformation of a pentapeptide purportedly constrained in an alpha-helical form by being tethered between its two ends with palladium. The analysis revealed that the alpha helix is not part of the ensemble of conformations that optimally fits the NMR data, with the dominant motif being a gamma turn<sup>113</sup>. More recently, structures from long, 100 ns MD simulations were combined with NMR data to settle a long-standing controversy about the involvement of gamma and  $\beta$  turns in the COOH-D-Pro-Ala-Ala-Ala-NH<sub>2</sub> pentapeptide. The joint computational-NMR analysis indicated that the peptide exists in two conformations in solution, and while both of them exhibit a  $\beta$ -II type turn, the gamma turn conformer is only a minor (7%) constituent of the solution<sup>114</sup>. These and other studies have demonstrated the value of using computational methods to disentangle time-averaged NMR data.

#### 4.2: The Stevastelins



**Figure 4.1: The Stevastelins**

Cyclic depsipeptides are cyclic peptides in which one or more of the amide nitrogens are replaced by oxygens, resulting in a lactone linkage. These depsipeptides are being pursued as anticancer, immunosuppressive and antimicrobial agents and one of them, kahalalide F, is in phase II clinical trials as a potential anticancer drug<sup>115</sup>. The stevastelins (**Figure 4.1**) are 13-membered cyclic depsipeptide natural products isolated from a culture broth of *Penicillium*<sup>116</sup>. In recent years they have been shown to have potent activity as immunomodulators and dual-specific phosphatase VHR inhibitors<sup>117</sup>. Relatively few compounds are known to inhibit this protein<sup>118,119</sup> and understanding of inhibitor action would benefit from biostructural investigations of known and new inhibitors.

Previous investigations have demonstrated the biological activity to depend on their configuration. To this end, a recent report by Bisek et al. described the synthesis, VHR-

inhibitory activity and conformational analysis of four stevastelin diastereomers and their phosphoserine analogs<sup>120</sup>. Based on force field conformational searches guided by NMR data obtained in DMSO-d<sub>6</sub>, the report concluded that each of the six stevastelin diastereomers essentially exists as a single family of conformations with variability only between the side-chain conformations and not in the backbone conformation. However, an alternative view suggests that the stevastelins with 9 rotatable bonds in their backbone may represent more conformational diversity than suggested in the recent analysis and may populate motifs in solution that are dissimilar from the ones proposed. To deconvolute the average NMR data and reveal the individual conformational components that contribute to it, we have combined the NMR data cited in the report with extensive multiple force field conformational searches and have performed NAMFIS analysis on the stevastelins. The NAMFIS analysis concludes that the conformational profile for the stevastelins can be better represented by a more variable conformational ensemble in solution which contains new motifs not found in the initial analysis<sup>121</sup>. We also supplement the NAMFIS analysis with ab initio quantum chemical calculations that indicate a different energetic hierarchy from that reported in the paper.

### ***Conformational searching***

For generating pertinent stevastelin structures we used the Macromodel program (v 9.6.1)<sup>122</sup> implemented in Maestro (Schrödinger, v 8.5.1. Extensive conformational searching for each of the four benzylated stevastelin diastereomers (**1**) was carried out with three force fields: AMBER\*, OPLS 2005, and MMFFs. Since NOE signals in the NMR data were restricted to C4 of the C3-alkyl side chain, the chain was excised at the

C4 carbon to simplify the torsional complexity in the searches. Each of the conformational searches used the Mixed Low Mode/Monte Carlo Method<sup>123</sup> with 30,000 steps, the GBSA/Water continuum solvation model<sup>124</sup>, a 7 kcal/mol energy window for saving the structures and a 0.25 Å all-atom RMSD criterion for structure elimination. The structures were optimized using the PRCG method with 500 steps per conformer. After the conformational search, the conformers from each calculation were further optimized to convergence with their respective force fields using the TNCG method with 100 steps per conformer. Finally the conformations from all three force fields were combined and duplicates were rejected with a 0.25 Å RMSD elimination criterion. The resulting pool of conformations was used as input for NAMFIS.

In addition, for comparison purposes, conformational searches with the OPLS 2005 force field constrained by the distances from the NMR data were run for all four diastereomers as per the protocol in the previous study<sup>120</sup>.

### ***NMR data***

The NMR data used was from the report by Bisek et al. The NOE peak intensities designated strong, medium and weak were converted into distance constraints of 2.3, 3.5 and 5.0 Å respectively as indicated in the paper. The  $^3J_{\text{H-N-C-H}}$  coupling constants were also used as stated in the report.

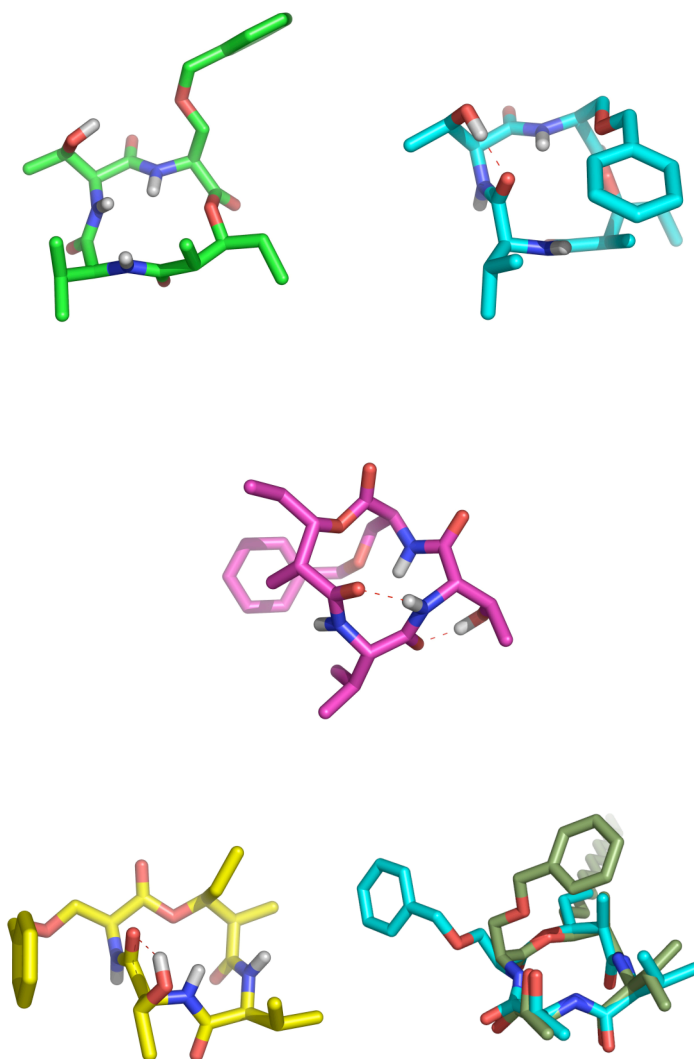


### ***4.3: Stevastelin conformations***

Since the original analysis systematically investigated the conformations of each individual diastereomer of **(1)**, we decided to adopt the same approach. To this end, NAMFIS analysis was performed on a set of conformations generated for each diastereomer. Below we document the results obtained for each of the diastereomers and discuss the similarities and differences we observed between these results and those described in the report by Bisek et al.<sup>120</sup> For comparison with the NAMFIS conformers, we used the corresponding top ten conformations for every diastereomer from the constrained conformational search protocol carried out in the previous study. We classify and compare similar structures based on backbone conformations.

**2R3S:** Conformational searching for this diastereomer with the listed protocol gave 3903 unique conformations from three force fields. NAMFIS analysis of these structures and the 2R3S NMR data generated 11 conformations ranging from 19-2% with a SSD (Sum of Square Difference) of 57. This value commonly falls below 200, and sometimes below 100<sup>4,5</sup>. However, the significance of the present value is underscored by the fact that similar NAMFIS processing of the previously proposed family of conformers alone gives a SSD of 103. The relatively much lower SSD of 57 establishes the stated goal of the NAMFIS analysis to produce multiple families of diverse conformations which fit the data better than any single family. This observation is underscored for all the other diastereomers analyzed below.

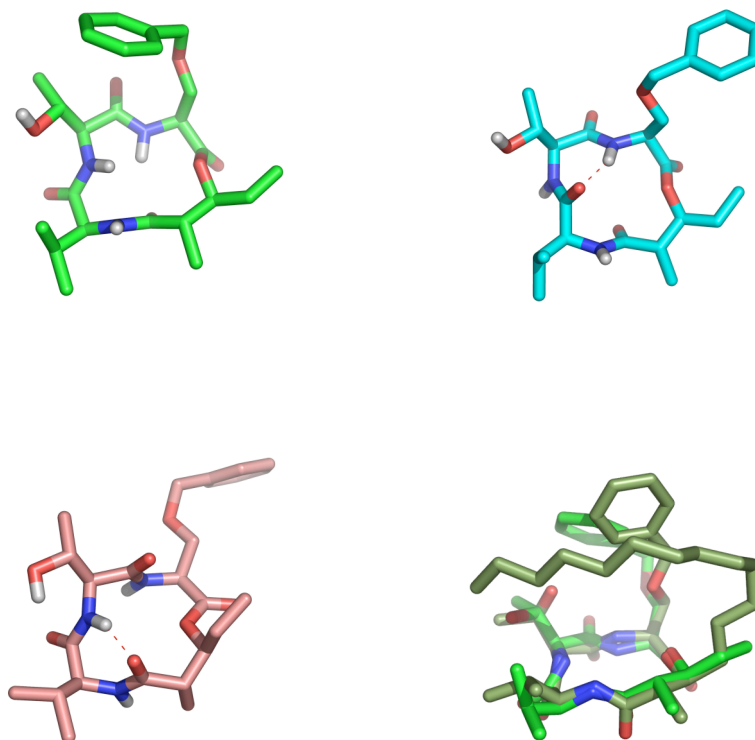
Based on clustering by visual inspection, the NAMFIS conformers can be classified into 3 families illustrated below (**Figure 4.2**). The largest family comprises of the 2<sup>nd</sup> (16%), 3<sup>rd</sup> (15%) and 4<sup>th</sup> (13%) best-fit conformers and contributes 44% to the conformational ensemble. This family posits a hydrogen bond between the Val carbonyl and the Thr side chain OH, which makes it dissimilar from the postulated dominant family. The second largest family contains the top (19%), 7<sup>th</sup> (7%) and 9<sup>th</sup> (4%) best-fit conformers and constitutes 39% of the population. This family is similar to the dominant family proposed in the previous paper and the two sets overlap with a mean backbone RMSD of 0.3 Å. There is a third minor family comprising of the 10<sup>th</sup> best fit (3%) with a hydrogen bond between the Thr OH and the Ser carbonyl. Another minor family contains the 8<sup>th</sup> best fit (5%) which has a gamma turn hydrogen bond between the C1 carbonyl and the NH. The classification of some of the conformers into families is more ambiguous. For example, the 6<sup>th</sup> best fit (7%) seems to belong to the dominant family (44%) but differs considerably in the O-C3-C2 region.



**Figure 4.2:** From left: Representative members of the four dominant NAMFIS families for 2R3S; best fit (19%), 2<sup>nd</sup> best fit (16%), 8<sup>th</sup> best fit (5%) and 10<sup>th</sup> best fit (3%), overlap of NAMFIS-1 and previously proposed conformation (olive, RMSD 0.3 Å)

Thus, the conformational analysis of the 2R3S diastereomer indicates that while some of NAMFIS the conformers resemble the earlier proposed conformers, the conformational ensemble is also populated by additional motifs different from the proposed one. For example the dominant family (44%) sustains a hydrogen bond unlike the proposed family, as do members of some minor families.

**2S3S:** Conformational searching for this diastereomer with the listed protocol gave 3708 unique conformations from three force fields. NAMFIS analysis of the data generated 9 conformations ranging from 18-4% with a SSD of 50. Based on clustering by backbone superposition and visual inspection, the conformers can be classified into 3 families illustrated below (**Figure 4.3**). The largest family contains the 2<sup>nd</sup> (16%), 3<sup>rd</sup> (16%) 4<sup>th</sup> (15%) and 9<sup>th</sup> (4%) best-fit conformations and contributes 51% to the conformational ensemble. Members of this family sustain a gamma turn hydrogen bond between the Val carbonyl and Ser NH. The second-largest family contributes 39% and comprises the best fit (18%), 5<sup>th</sup> (10%), 7<sup>th</sup> (7%) and 8<sup>th</sup> (4%) best-fit conformations. This family resembles the dominant conformational family and the two sets overlap with a backbone RMSD of 0.4 Å. In addition there is a third minor family which consists of the 6<sup>th</sup> best fit (10%) conformer; this conformer has a gamma turn hydrogen bond between the C1 carbonyl and Thr NH.



**Figure 4.3:** From left: Representative members of the three dominant NAMFIS families for 2S3S; best fit (18%), 2<sup>nd</sup> best fit (16%) and 6<sup>th</sup> best fit (10%), overlap of NAMFIS-1 and previously proposed conformation (olive, RMSD 0.3 Å)

Comparison of the NAMFIS conformer sets for 2S3S with the 2R3S dataset above illuminates similar trends. As in the case of 2R3S, there are diverse conformer families that significantly populate the conformational ensemble, with both the dominant family (47%) and the minor family (10%) containing a gamma turn hydrogen bond.

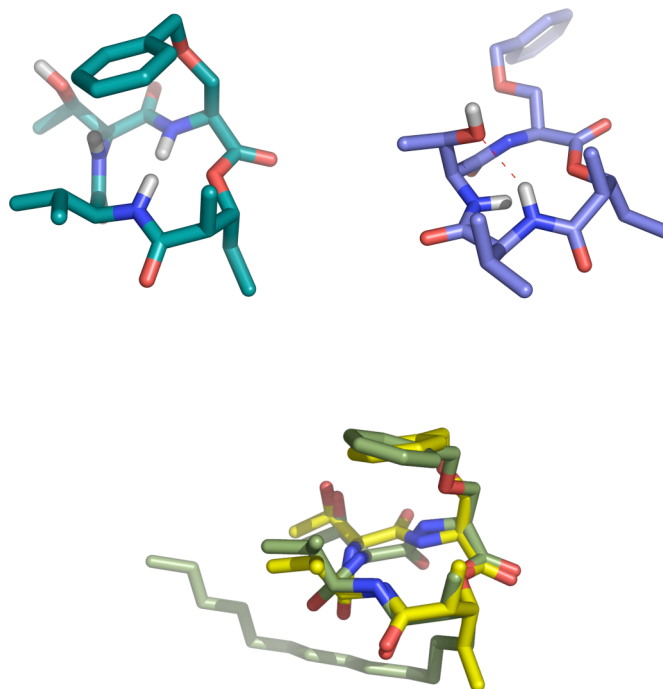
Comparison of the 2R3S and 2S3S NAMFIS conformations also indicates close similarities as well as some differences between the two sets. For instance both

diastereomers contribute dominant conformational families that are present to similar extents (44% and 47%) except that one family (2R3S) contains a Val carbonyl-Thr OH hydrogen bond while the other (2S3S) contains a Val carbonyl-Ser NH gamma turn hydrogen bond. Both these families also overlap closely in their backbone conformations (mean backbone RMSD 0.3 Å). Similarly, the second-largest families also closely overlap in their backbone conformations (RMSD 0.4 Å), populate the solution ensemble to the same extent (39%) and contain no hydrogen bonds. Interestingly the rank of the conformers that contribute to these similarly-populated families are also the same for both 2R3S and 2S3S; for instance it's the 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> best fit conformations that contribute to the dominant family in each case. Apart from these resemblances, the remaining conformers are different in the two datasets indicating conformational variability.

Thus, the observation in the paper of common conformations for 2R3S and 2S3S is partly justified but also masks conformational heterogeneity. While there are indeed common conformations present to similar extents, they differ in hydrogen bonding characteristics and in addition, both diastereomers contribute unique conformations different from the one postulated.

**2R3R:** Conformational searching for this diastereomer with the listed protocol gave 2600 unique conformations from three force fields. NAMFIS analysis of the data generated 6 conformations ranging from 44-4% with a SSD of 62. Based on clustering by visual inspection, the conformers can be classified into 3 families illustrated below (**Figure 4.4**).

In this case unlike before, there is one significant conformational family populating the conformational ensemble. The dominant member of this family is the best fit (44%) conformer, but the 2<sup>nd</sup> (21%), 3<sup>rd</sup> (17%) and 6<sup>th</sup> (4%) best fit conformations also have very similar backbone conformations (average RMSD 0.3 Å) and are therefore part of the same family. This family contributes 86% and hence represents almost the entirety of the conformational ensemble. Concomitantly it also resembles the dominant conformer family previously postulated, with minor differences in the backbone of the two sets resulting in a RMSD of 0.5 Å. However, the NAMFIS conformers also include an additional dissimilar family exemplified by the 4<sup>th</sup> (9%) and 5<sup>th</sup> (6%) best fit conformers; together these contribute 15% and represent a non-trivial fraction of the solution population. This second family is characterized by a Val NH-Thr OH hydrogen bond.

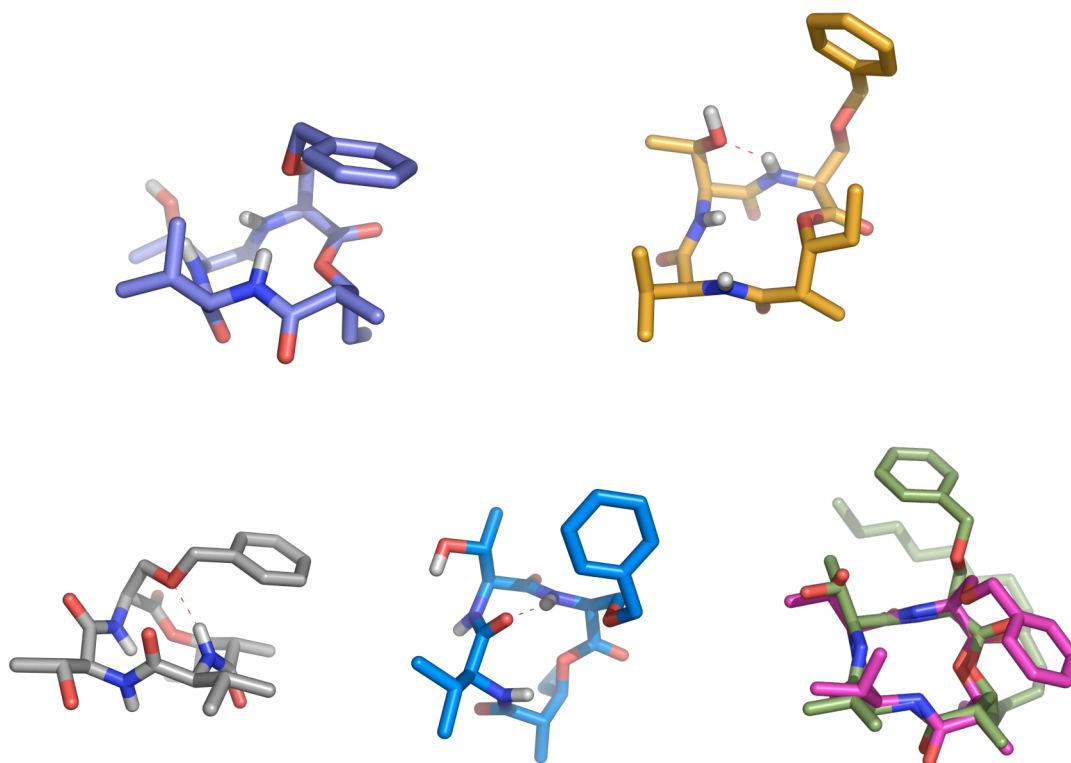


**Figure 4.4:** Representatives of the two NAMFIS families for 2R3R. Left: Best-fit (44%) contributing to the dominant family (86%). Right: 4<sup>th</sup> best fit (9%) contributing to the second-most dominant family (15%). Bottom: Overlap of NAMFIS-1 with previously proposed conformer (RMSD 0.5 Å)

**2S3R:** Conformational searching for this diastereomer with the listed protocol gave 2799 unique conformations from three force fields. NAMFIS analysis of the data generated 5 conformations ranging from 60-6% with a SSD of 62. Clustering of these conformations by superposition indicates four families (**Figure 4.5**). The best fit (60%) and 3<sup>rd</sup> best fit (8%) conformers are identical in their backbone conformations and thus again represent one dominant family contributing 68% to the solution ensemble. This family resembles



the postulated dominant family in its lack of hydrogen bonds and the two sets overlap with a backbone RMSD of 0.4 Å. The three remaining NAMFIS conformations are all different and constitute their own families. The 2<sup>nd</sup> best fit (18%) has a hydrogen bond between the Thr OH and the Ser NH. The 4<sup>th</sup> best fit (8%) conformer sustains a hydrogen bond between the benzyloxy oxygen and the Val NH. The 5<sup>th</sup> best fit (6%) has a gamma turn hydrogen bond between the Val carbonyl and Ser NH (**Figure 4.5**). Thus, while the 2S3R diastereomer like the 2R3R diastereomer is represented by one dominant solution conformation, there is still a wide range of dissimilar conformations that contribute measurable fractions to the solution ensemble.



**Figure 4.5:** From left: Representative members of the four major NAMFIS families for 2S3R. From left; best fit (60%), 2<sup>nd</sup> best fit (18%), 4<sup>th</sup> best fit (8%) and 5<sup>th</sup> best fit (6%), overlap of NAMFIS-1 and previously proposed conformation (olive, RMSD 0.4 Å)

The above analysis supports one of the significant conclusions in the previous paper, namely that a change in the stereochemistry at C3 from S to R leads to a dramatic difference in the solution conformational ensemble. Thus in light of this fact, none of the backbone conformations in the 3S dataset are similar to the ones in the 3R dataset and both sets of conformations display divergent hydrogen bonding patterns. At the same

time, there is still conformational heterogeneity in both datasets that can be best represented by multiple conformations- some of them constituting significantly dissimilar families- that fit the NMR data. This conformational heterogeneity is demonstrated for example by the presence of conformers with gamma turns in many of the datasets.

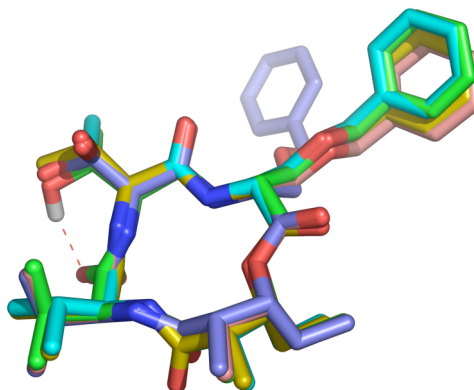
#### ***4.4: Effect of d-Serine on conformations***

To investigate the effect of a d-Serine residue on the stevastelin conformations, we also did a NAMFIS analysis of the 2R3S and 2S3R d-Ser diastereomers. As before, we particularly wanted to analyze the effects of the d-Ser residue on conformational heterogeneity and energetic hierarchy.

##### ***2S3R d-Ser:***

Conformational searching for this diastereomer with the listed protocol gave 3500 unique conformations from three force fields. NAMFIS analysis of the data generated 5 conformations ranging from 71-2% with a SSD of 51. In this case, the best fit (71%), 2<sup>nd</sup> (15%) 3<sup>rd</sup> (7%), 4<sup>th</sup> (6%) and 5<sup>th</sup> (2%) best-fit conformers are all not only very similar in their backbone conformations (mean RMSD 0.5 Å) but four of them (1, 3, 4 and 5) are also similar in their side chain conformations (**Figure 4.6**). Four of these conformations are not hydrogen bonded, although in the 4<sup>th</sup> best-fit conformer, the slightly different orientation of the Thr OH makes a hydrogen bond between this functionality and the Val carbonyl possible. Thus the 2S3R d-Ser diastereomer emerges as the first and only example in this study that truly exists as a single, dominant conformational family in solution. Furthermore, since NAMFIS also locates a single dominant conformer family

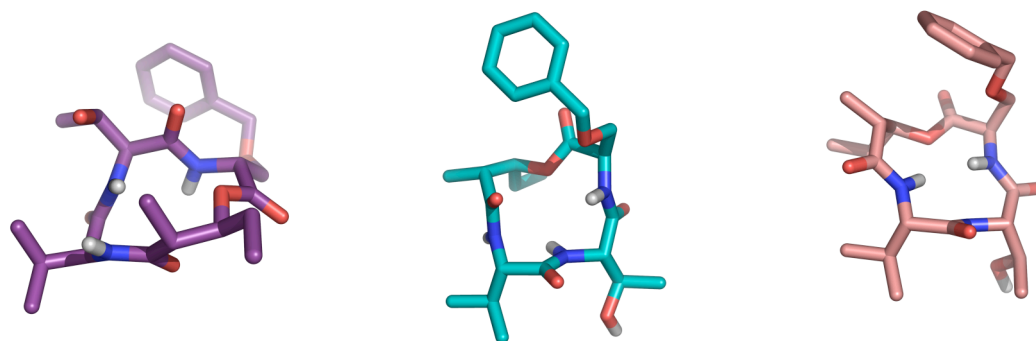
for this diastereomer, this family closely resembles the corresponding postulated family in the previous study, except in the disposition of the benzyl side chains.



**Figure 4.6:** Superposition of the five NAMFIS conformers for 2S3R d-Ser, with the 4<sup>th</sup> best fit (green) sustaining a hydrogen bond between the Val carbonyl and Thr OH

**2R3S D-Ser:** Finally, conformational searching for this diastereomer with the listed protocol gave 4608 unique conformations from three force fields. NAMFIS analysis of the data generated 6 conformations ranging from 52-2% with a SSD of 62. Clustering of the conformations by superposition indicates that the best fit (52%), 2<sup>nd</sup> (21%) and 3<sup>rd</sup> (17%) best fit conformations are virtually identical in their backbone conformations (RMSD 0.3 Å). Thus they constitute the dominant conformational family with a contribution of 90% and, similar to the 2S3R d-Ser NAMFIS conformer, closely resemble the conformation postulated in the paper. The top two conformations do not have hydrogen bonds while the 3<sup>rd</sup> best fit conformation projects a slightly different orientation of the Thr OH and therefore sustains a hydrogen bond between this functionality and the Val carbonyl. The 4<sup>th</sup> (5%) and 5<sup>th</sup> (3%) best fit conformations are

also very similar and constitute one family, with the 5<sup>th</sup> best fit again sustaining a hydrogen bond between the Thr OH and Thr carbonyl. The 6<sup>th</sup> best fit (2%) is different from all others and does not sustain any hydrogen bonds. **Figure 4.7** illustrates the three NAMFIS families. Thus the 2R3S d-Ser diastereomer is characterized by a dominant (90%) conformational family in solution, but with measurable contributions (10%) from two other families.



**Figure 4.7:** Representative conformations from the three NAMFIS families for 2R3S d-Ser. From left: best fit (52%), 4<sup>th</sup> best fit (5%) and 6<sup>th</sup> best fit (2%)

#### ***4.5: Energetic analysis:***

The previous report also includes an energetic analysis of the four diastereomers that ranks their proposed dominant conformations based on OPLS 2005 force field energies. However, force field energies are widely subject to variability because of different parametrization and convergence criteria and are frequently overemphasized by the dominance of electrostatic interactions. This fact renders energetic ranking of conformers using force fields energies complicated at best, and clearly suggests a strong measure of

caution in trusting energetic rankings from a single force field. For instance, global minima identified by one force field often may not correspond to those identified by others<sup>32</sup>. In the original report the conformers whose energies are compared are those that are obtained as low-energy structures from an OPLS 2005 conformational search guided by NMR constraints. However, a more prudent energetic comparison would be between the solution experimental global minima deduced by NAMFIS, that is, the best fit conformer for each diastereomer

To this end, quantum chemical *ab initio* calculations were carried out with the Jaguar program (v 7.5.1) implemented in Maestro (Schrödinger, v 8.5.1). The *ab initio* calculations were done at the B3LYP/6-31G\*\* level with the PBF-DMSO d<sub>6</sub> continuum solvation model<sup>125</sup>. All calculations were single-point energetic calculations on constrained-optimized OPLS 2005 structures. Since the NAMFIS conformers come from a conformational pool that includes local energy minima from three different force fields, prior to the *ab initio* calculations each conformer was optimized with atomic position constraints in the OPLS 2005 force field to ensure that all structures are on the same potential energy surface as the conformers in the previous report. For the MMFFs comparisons the conformers were constrained-optimized with the MMFFs force field. Atomic constraints with a force constant of 100 kJ/mol/Å<sup>2</sup> were used for the force field optimization.

The *ab initio* and force field calculations indicate a different energetic ranking than that for the low-energy structures in the paper (**Table 4.1**).

NAMFIS Glob. Min. Diastereomer	$\Delta E_{\text{Rel.}}$ (kcal/mol) (OPLS-05)	$\Delta E_{\text{Rel.}}$ (kcal/mol) (MMFFs)	$\Delta E_{\text{Rel.}}$ (kcal/mol) (B3LYP/6-31G**//PBF, DMSO)
2R3S	0.0	13.1	3.9
2S3R	7.0	0.0	3.0
2R3R	7.6	7.8	9.1
2S3S	9.4	12.8	10.1
2S3R-d-Ser	4.5	0.6	11.0
2R3S- d- Ser	7.6	10.8	0.0

**Table 4.1:** Relative force field and ab initio energies for the six NAMFIS global minimum diastereomers

According to the previous analysis, the energetic ranking of the four OPLS 2005 global minima was 2S3S < 2S3R ~ 2R3S < 2R3R within a 7 kcal/mol window. However, the quantum chemical and force field calculations on the solution global minima indicate a different ranking; while the three methods give divergent results for the energetic hierarchy of the stereoisomers, all three indicate the 2S3S to be much higher in energy as shown in Table 1. The calculations also suggest a differing energetic ranking for the two d-Ser conformers. While the force fields locate the 2S3R d-Ser conformer to be intermediate in relative energy, the ab initio calculation posits it as the highest energy conformation, while also indicating its 2R3S diastereomer to be the most energetically stable. These conclusions, although not supposed to indicate a quantitative energetic

ranking, differ from those in the previous study which suggest both d-Ser conformers to be higher in energy than their d-Ser counterparts.

#### ***4.6: Conclusions and outlook***

To summarize then, we performed a conformational deconvolution of the average NMR spectra for six stevastelin diastereomers using NAMFIS and have derived novel conformational motifs not observed earlier, including conformers with gamma turns. NAMFIS analysis also confirms that a change in chirality at the C3 center leads to a dramatic change in conformations as postulated in the paper although some of the resulting conformers are different from the previous ones.

The analysis suggests that cyclic peptides, even with their relatively constrained configurations, can exhibit conformational variability resulting from easily rotatable bonds in their backbone. At the same time, partial rigidity enforced by cyclization is exemplified by the fact that the NAMFIS results do include some conformers similar to the previously proposed ones, including the top conformers in certain datasets. In the past, cyclization has been shown to be an effective strategy for maintaining conformational homogeneity<sup>126,127</sup>. The present study indicates that while cyclization is indeed a relevant strategy for enforcing rigidity, it may also mask the presence of diverse conformations that may possibly contribute to the biological activity of the molecule. This observation underscores the general principle that constraining elements in peptide design can nonetheless impart a degree of flexibility that does not readily reveal itself in averaged NMR data.



At the same time in the current study, the dominant conformations deduced by NAMFIS for the two d-Ser diastereomers are exceptions and closely resemble the conformers postulated earlier. Thus, d-Ser emerges as a conformational constraint that reduces conformational mobility in the stevastelins. This fact could be used as a design element that could force the stevastelins and similar molecules to adopt a specific dominant conformation in solution.

In the present study, NAMFIS analysis of the phosphorylated stevastelin analogs could not be attempted due to lack of published NMR data on these molecules. Such analysis could be valuable in shedding light on the differences in conformation between the phosphorylated and unphosphorylated diastereomers. In the previous report it was hypothesized that the phosphorylated analogs possess intrinsically higher flexibility. An alternative hypothesis arising from this study suggests that rather than inducing greater conformational flexibility, phosphorylation could be stabilizing a particular conformation of the unphosphorylated analogs and this could be the bioactive conformation. Thus a similar study of the phosphorylated analogs would importantly indicate whether there is a dramatic change in the conformational ensemble upon phosphorylation, or whether phosphorylation simply entails a change in the population distribution of already existing conformer families.

In previous studies on bioactive small molecules binding to proteins, the bioactive, protein-bound conformation for flexible molecules has been located as a minor constituent in solution<sup>33,36</sup>. Since structural information about the bioactive conformation

of the stevastelins is lacking, the above conformational deconvolution usefully derives a set of low-energy solution conformations, one of which as stated above may be the bioactive one. This data may prove useful in the understanding of the binding mode of the stevastelins and the design of improved analogs.

# **Chapter 5**

## **Selective Inhibitors of Cyclin-Dependent Kinase 7**

“Luck is what happens when preparation meets opportunity”

-Seneca the Younger

### **5.1: Cancer**

**E**ven after decades of massive spending on research and care, cancer is still one of the world's biggest killers. As of 2007, it afflicted about 10.9 million people worldwide and caused the deaths of 6.7 million<sup>128</sup>. More research resources and funding have been devoted to cancer than perhaps to any other medical problem. Since 1971, the United States has invested over 200 billion dollars in cancer research, and yet in 2008 cancer was responsible for 1.4 million new cases and 565,000 deaths in the country, or about 1,500 people a day<sup>129</sup>. This makes it the second-largest killer after heart disease in the United States.

Several decades of research have shed light on the causes and mechanisms of cancer. The generally accepted view posits a breakdown in the genetic machinery of a particular cell somewhere in the body<sup>130,131</sup>. This breakdown allows the cell to bypass many of the normal checkpoints that prevent excessive cell growth and proliferation<sup>130</sup>. Once the checkpoints are circumvented, the cell rapidly and uncontrollably divides. At some point, its progeny dissociate from the parent cell and migrate to other organs and parts of the body through the body's vascular and lymphatic network. This well-known process, called metastasis, establishes populations of cancerous cells in other parts of the body that gradually but surely start interfering with normal functions of other cells, organelles, organs and systems. Once metastasis begins and progresses, mortality rapidly increases.

Cancer research has also considerably invested in understanding the causes of cancer. This research has shown that 'cancer' is not a single disease but a manifestation of

complex factors and mechanisms leading to metastasis. To date, many different factors including genetic, environmental, and lifestyle factors have been discovered to be important in carcinogenesis although the exact etiology can be difficult to determine in individual cases. The connection between cigarette smoking and lung cancer for instance is well established, but the role of dietary constituents in contributing to cancer is much less known. The discovery of oncogenes and tumor-suppressing genes has brought fundamental insight into the molecular mechanisms of cancer<sup>131</sup>. The frequently-mutated p53 gene TP53 for example is now widely known to be central to many of the cell's anti-cancer mechanisms, including promotion of apoptosis and DNA repair<sup>132</sup>. In investigating the causes of cancer, focus has been particularly given to genes and their resultant protein molecules that are overexpressed and that lead to uncontrolled cell division. Sometimes a single gene, gene family or protein family has been discovered to be the cause or risk factor in a particular type of cancer. For example, the Bcr-Abl kinase protein is overexpressed in chronic myeloid leukemia and results from the constitutively expressed fusion gene that's created by the translocation of parts of two different chromosomes<sup>133</sup>. Similarly, mutations in the BRCA1 and BRCA2 genes have been identified as risk factors for breast cancer<sup>134</sup>. Another key protein overexpressed in up to 30% breast cancers is the tyrosine kinase receptor HER2. A recent antibody that targets this receptor called Herceptin has emerged as a breakthrough treatment in breast cancer<sup>135</sup>. While such single proteins or mutations are sometimes discovered as responsible for certain tumors, more often multiple and complex causes lead to carcinogenesis, consequently making it difficult to target cancer with chemotherapy.

The traditional three-pronged approach towards treating cancer has consisted of radiation, surgery and chemotherapy. Both radiation and surgery can be debilitating and while their use will remain valuable, chemotherapy seems to be the most evolving strategy towards targeting the disease. Due to the complex nature of cancer and the lack of a general cure for it, the search for therapies that will treat or cure particular forms of cancer has garnered intense attention for more than a century<sup>136</sup>. The treatment of cancer is an inherently complex task since cancerous cells frequently exhibit characteristics which are similar to those of normal cells, even if they express these characteristics aberrantly. For instance, many proteins that result in cancer cell proliferation are also present in normal cells, albeit expressed at low levels. Because of this fact and the urgent nature of the treatment needed for reducing mortality, many past and existing therapies have had relatively low concern for side effects, mainly focusing on eliminating cancerous cells. Thus, cytotoxic agents have long played an important role in cancer therapy<sup>136</sup>. Among the earliest such agents were the sulfur and nitrogen mustards; the toxic qualities of these compounds are apparent from the fact that some of them such as mustard gas were used as chemical warfare agents in World War I<sup>137</sup>. Later key cancer therapeutics such as cisplatin<sup>138</sup> and DNA-intercalating agents<sup>139</sup> were more potent and selective but still retained the toxic side effects of other compounds. More often than not, cancer chemotherapy's widespread use stems simply from the lack of knowledge of more selective and safer compounds, and the sheer need to effect quick treatment in the face of rapid and high mortality. Thus, while many potent compounds exist for treating cancer and while many more will be discovered, safety still remains a priority and an unmet general need in cancer therapy.

Because cancer arises from multiple and complex signaling mechanisms that cells use to circumvent natural inhibitory checkpoints, there is considerably network redundancy built into these mechanisms, and this fact frequently thwarts cancer therapy wherein the subjugation of one protein involved in a signaling pathway will simply be compensated for by the overexpression of another protein in another pathway<sup>140</sup>. In addition, molecular targets in cancer also demonstrate the common mechanism of developing resistance to therapy through mutations in their binding sites. Cancer cells can also render drugs less effective through expression of efflux pump proteins that extrude cytotoxic compounds out of the cell. For instance, resistance through taxol has been shown to develop through both these mechanisms<sup>141</sup>. Due to this mechanistic redundancy and development of resistance, it becomes necessary, and is indeed common, to administer multiple chemotherapeutic agents to target different pathways. However, these modes of resistance also make it necessary to investigate both modified analogues of existing drugs that would be immune to resistance, as well as new molecular targets which can be inhibited by current and new therapies. To this end, there are major research programs directed into finding novel protein targets involved in cancer. In Chapter 2 we have already seen the value of drugs that target microtubule assembly in cell division. Another very important family of proteins that have recently been found to be key targets in cancer is the protein kinase family.

## ***5.2: Phosphorylation***

The essence of complex life is signaling. Almost every key task that life carries out involves communication between molecules, cells, organelles and systems. Over billions of years, evolution has produced extremely sophisticated shared mechanisms for mediating signaling between cells<sup>142</sup>. In complex eukaryotic organisms, one of the most common communication mechanisms involves the activation of membrane proteins by small molecules. The signal is then further amplified by a series of steps sometimes involving chemical modifications of dozens of other small molecules, peptides and proteins whose endpoint can be a key process like gene expression<sup>142,143</sup>. For instance, hormones in organisms often bind to nuclear receptor proteins which then oligomerize and bind to stretches of DNA, thus controlling DNA expression and cell proliferation and regulation<sup>144</sup>. Other well known examples include the adrenaline receptor that is activated by the small molecule adrenaline and growth factor receptors that can be activated by small molecules or peptides, resulting in cell growth and proliferation.

Among all the signaling mechanisms prevalent in living organisms, phosphorylation occupies a central place as the single-most important and widespread mechanism of cellular communication<sup>145,146</sup>. The phosphate group is a small, polar chemical moiety which can have profound effects when it is attached to molecules and proteins. Phosphate linkages constitute the key connecting links in DNA and RNA and triphosphates in the form of ATP are the ‘currency of life’, the major energy-providing machinery in the cell. In a seminal article, Westheimer described the special properties of phosphate groups as emerging from their fundamental acid-base chemistry<sup>147</sup>. Westheimer compared the



phosphate group to many other possible chemical groups that could possibly have been chosen by life as key molecular elements, and concluded that the specific ionization and pKa characteristics of phosphate groups confer the right balance of stability and susceptibility to hydrolysis necessary for carrying out important functions in the cell. For instance, the unique pKa value of the phosphate group ensures that it is doubly ionized and thus protects phosphodiester linkages in nucleic acids from undergoing rapid hydrolysis from nucleophilic attack by water. On the other hand, the delocalization of charge in triphosphates also renders their hydrolysis in ATP enormously energetically favorable; a key necessity for the energetic coupling that ATP performs in facilitating otherwise energetically unfavorable reactions.

### ***5.3: Protein kinases***

Due to these special properties of phosphates, phosphorylation has also been chosen by nature as the dominant mechanism for transmission of genetic information and for signal transduction<sup>4</sup>. For signaling purposes nature has evolved a class of proteins called protein kinases that phosphorylate substrates<sup>148</sup>. These substrates are usually other proteins and peptides. Protein kinases themselves can get phosphorylated by other kinases.

Phosphorylation of and by kinases always occurs at a hydroxyl residue of one of the three hydroxyl-containing amino acids; threonine, serine or tyrosine<sup>149</sup>. Protein kinases have been divided into two classes depending on whether they phosphorylate the aromatic residue (Tyr) or the aliphatic residues (Ser and Thr). Tyrosine kinases are further divided into receptor-tyrosine kinases (RTKs) and non-receptor tyrosine kinases.<sup>150</sup>

Ever since phosphorylation by protein kinases was discovered as a dominant signaling mechanism, more than 500 different kinases playing a remarkable variety of roles have been discovered in eukaryotes<sup>151</sup>. Kinases have been implicated in central signal transduction processes whose effects range from cell growth to regulation of hormonal action and brain and cardiac function. Based on genome studies, 518 kinases have been documented to be a part of the protein ‘kinome’<sup>151</sup>. Several of these kinases are expressed as multiple isoforms which have differential tissue distribution and subtly different roles in signal transduction pathways.

In recent years, kinases have become a primary focus in drug discovery. Kinases have become important targets for drugs because of the observation that many kinases serve as key regulators of cell growth, differentiation and proliferation and are constitutively expressed or deregulated in diseases such as cancer. Important kinases involved in cell growth and proliferation include epidermal growth factor receptor (EGFR)<sup>152</sup>, platelet-derived growth factor receptor (PDGFR), mitogen-activated kinase (MAP)<sup>153</sup> and phosphoinositide-3-kinase (PI3K)<sup>154,155</sup>. These and other kinases are crucial components of the cell’s signaling machinery. Some such as EGFR are membrane receptor kinases that bind to extracellular ligands or peptides, oligomerize and then activate other kinases<sup>156</sup>. These kinases in turn activate other biomolecules in cascade-like events and cause an amplification of the initial signal, whose end result is usually DNA regulation, growth and proliferation. The activation of PI3K for example ultimately results in the activation of the key transcription factor NF- $\kappa$ B<sup>154</sup>.

#### ***5.4: Kinase inhibitors in medicine***

Over the last decade, the treatment of cancer has seen a new paradigm with the introduction of kinase inhibitors, and kinase-targeting drugs and antibodies for treating various types of cancer are already in clinical trials and on the market<sup>150,157,158</sup> (**Table 5.1**). Kinase inhibitors for treating inflammation, diabetes, heart disease and Alzheimer's disease are also under serious investigation although progress in these areas has not been as prevalent as that in oncology.

U.S. Brand name	Year approved	Generic name	F.D.A approved indications	Company	Target kinases
Gleevec	2001	Imatinib	Chronic myeloid leukemia (CML)	Novartis	Abl, c-Kit, PDGFRa, PDGFRb
Iressa	2003	Gefitinib	Non-small-cell lung cancers (NSCLC)	AstraZeneca	EGFR
Tarceva	2004	Erlotinib	NSCLC, pancreatic cancers	Genentech, OSIP	EGFR
Nexavir	2005	Sorafenib	Hepatocellular carcinoma, renal cell carcinoma	Bayer, Onyx	Raf, VEGFR2, VEGFR3, c-Kit, PDGFRb
Sutent	2006	Sunitinib	Gastrointestinal stromal tumor (GIST), renal cell carcinoma (RCC)	Pfizer	c-Kit, VEGFR, PDGFR, FLT3
Sprycel	2006	Dasatinib	CML	Bristol-Myers Squibb	Abl, c-Kit, PDGFR, Src
Tasigna	2007	Nilotinib	CML	Novartis	Abl, c-Kit, PDGFRb, Src, Ephthrin
Tykerb	2007	Labatinib	Breast cancer	GlaxoSmithKline	EGFR, Her-2

**Table 5.1:** List of approved anticancer kinase inhibitors currently on the market (Table from Bikker et al.<sup>159</sup>)

It is only in recent years that the validation of kinases as therapeutic targets has emerged as a viable paradigm in drug discovery. Initial skepticism seemed justified by the lack of selectivity that was predicted to result from the very similar ATP-binding sites that kinase inhibitors would target. Since toxicity usually results from off-target effects due to multitarget inhibition, it was not clear how kinase-targeting drugs would become valuable therapeutic agents. However, after several years of research culminating in the approval of bonafide kinase inhibitors, it has become clear that subtle differences in the ATP binding site as well as the possibility of targeting dissimilar allosteric sites can make it possible to discover selective kinase inhibitors to treat diseases like cancer with a reasonable margin of safety<sup>160</sup>.

As the paradigm of kinase targeting has advanced, so has the debate between selective single-kinase and multiple-kinase inhibitors. At the onset it was reasonably surmised that potent single kinase inhibitors would provide the greatest degree of safety due to minimum off-target effects. However, some well-known kinase inhibitors that were initially designed to be selective agents were later discovered to target other kinases. In fact, among the dozen or so kinase inhibitors currently on the market, only one or two are now thought to exercise their effects through single kinase inhibition (**Table 5.1**), an observation that may be revised by future studies<sup>159</sup>. In probably the best-known case, the kinase inhibitor imatinib (Gleevec®, Novartis) which was designed to target the aberrantly regulated Bcr-Abl kinase and treat chronic myeloid leukemia (CML) was found to also exert its effects on two other kinases, c-Kit and PDGFR, albeit not as potently as on Bcr-Abl<sup>133</sup>. The Bcr-Abl kinase is produced through the expression of a

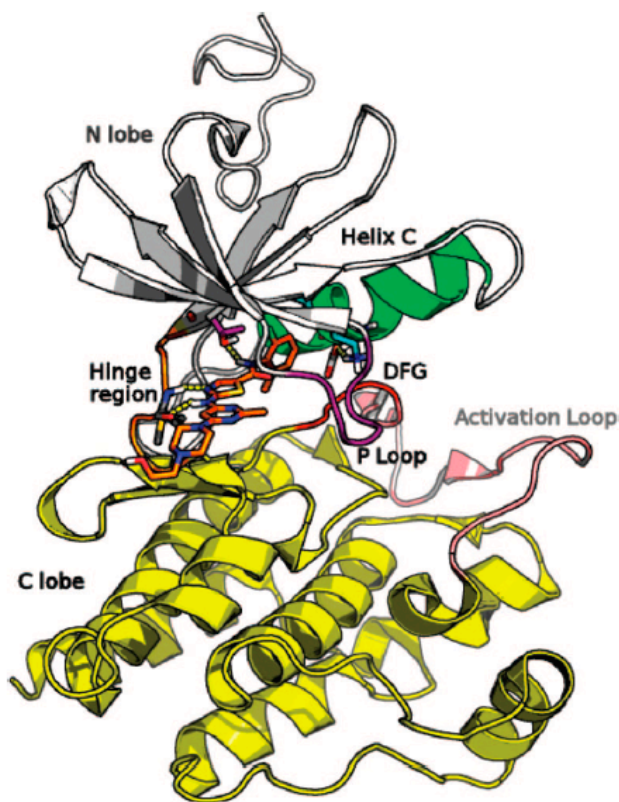
unique mutant chromosomal translocation and therefore is expressed only in cancer cells. Thus its targeting was thought to be a close-to-ideal scenario for cancer therapy. However, imatinib's role as an effective anti-cancer drug was clearly enhanced by inhibition of the other two kinases. Interestingly its inhibition of these kinases made it possible to utilize it in the treatment of another type of malignancy in which these targets are upregulated, namely gastrointestinal stromal tumours<sup>161</sup>. In some cases, multiple targeting clearly provides benefit by debilitating multiple pathways recruited by cancer cells for proliferation and survival; for example the kinase inhibitor sunitinib (Sutent®, Pfizer) acts primarily by targeting the c-Kit kinase constitutively expressed in tumors, but it also exercises a secondary effect by inhibiting VEGFR and PDGFR kinases which are crucial components of the cancer cell's angiogenesis machinery<sup>162</sup>.

Thus, the virtues of single versus multiple kinase inhibitors constitute ongoing research. In this context, the author participated in a "Single vs. Multiple Kinase Inhibitors" discussion in a 2008 kinase conference and the general consensus seemed to be that the best outcome depends on the specific kind of cancer, knowledge of the operative biochemical pathways and the patient population. The question is rendered complicated by the complex nature of the pathways that control signaling in which ideally inhibiting only a select subset of kinases may be most effective for anti-cancer therapy. However, such kind of target validation constitutes a major scientific endeavor both for specific kinds of tumors and also because of patient variability that may thwart standardized treatment. In addition, while there are several pan-kinase inhibitors such as staurosporine known, these compounds target many kinases indiscriminately and therefore are of

limited or no use in treatment. Designing both single kinase inhibitors and inhibitors targeting a judiciously chosen subset of well-validated kinases seems to be the major challenges in future kinase design, although some progress has been seen recently<sup>163</sup>.

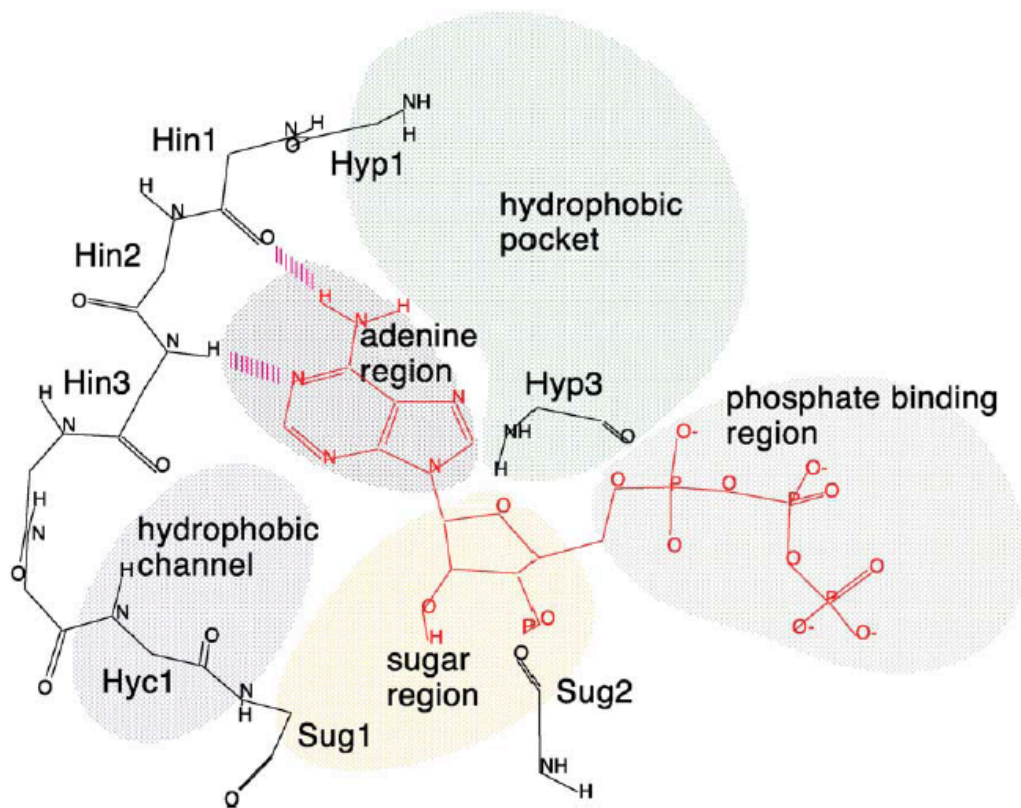
### 5.5: Kinase structure

All kinases share a common catalytic superstructure consisting of an N-terminal domain composed of  $\beta$  sheets with a single  $\alpha$  helix (helix C) and a C-terminal domain composed of  $\alpha$  helices (**Figure 5.1**).



**Figure 5.1:** The inhibitor dasatinib bound to Abl kinase illustrating the key features of kinase structure (Figure from Bikker et al.<sup>159</sup>)

A short strand called the hinge region connects the two lobes. The ATP-binding site of kinases consists of a well-defined pocket sandwiched between the two domains adjacent to the hinge region. ATP engages in critical hydrogen bonding and non-polar interactions with residues in the hinge region and elsewhere (**Figure 5.2**).



**Figure 5.2:** Interactions of ATP with residues in the hinge region and binding pocket of a generic kinase (Figure from Fabbro et al.<sup>164</sup>)

All kinase inhibitors without exception also interact with the hinge region. The ATP-binding sites of many kinases share close sequence homology<sup>160</sup>. A loop called the P-loop



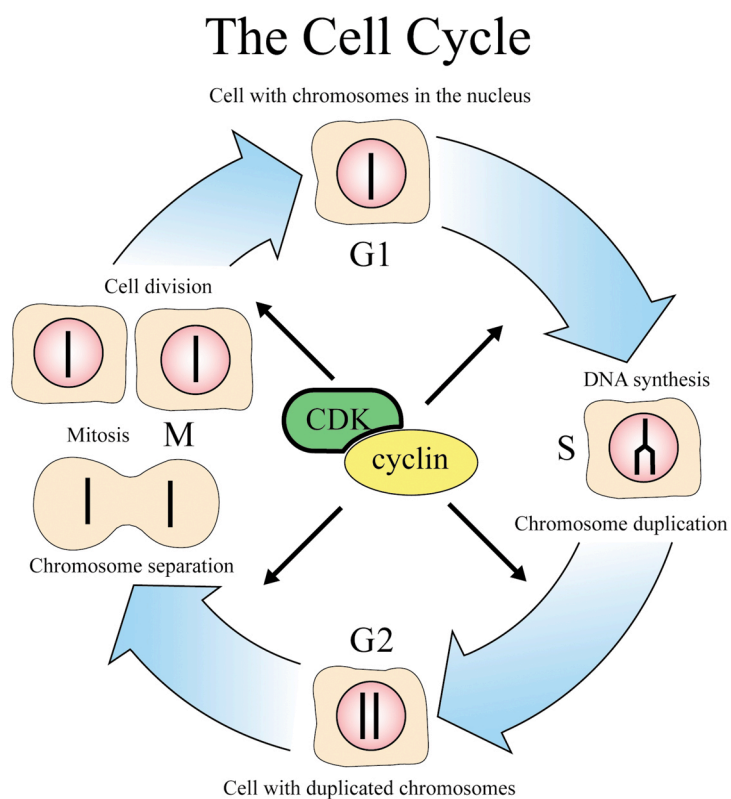
with the sequence G-X-G-X-X-G plays a large role in the dynamics of the kinase domain and controls the shape of the ATP-binding pocket; another loop called the activation loop can get phosphorylated by other kinases and is crucial for catalytic activity (**Figure 5.1**). Kinases usually exist in two states; an inactive state which does not bind ATP and an active state in which ATP binds and transfers a phosphate group to the substrate. In many pharmacologically important kinases the transition between the active and inactive states is modulated by the key DFG motif which exists in two states; in the DFG-out state, the Phe in the motif blocks the ATP-binding site and prevents activation while in the DFG-in state the motif swings away from the active site and allows activation. Some kinase inhibitors have been designed to interact with this motif to prevent activation. In addition, there is a crucial residue called the gatekeeper residue adjacent to the hinge residues which controls selectivity in the binding of ligands<sup>160,165</sup>. Modulating the shape and size of functional groups on inhibitors that target the gatekeeper can modulate kinase selectivity.

All kinase inhibitors have to address two important selectivity and potency problems that are interrelated. As noted above, the first problem with selectivity arises from very similar ATP binding sites. Secondly, inhibitor binding has to compete with ATP binding and since ATP concentrations in the cell -typically about 1-5 mM- are much higher than inhibitor concentrations, inhibitor potency has to surpass a certain value before the inhibitor can fruitfully exert its action on the kinase<sup>166</sup>. Minimum potency is also paramount for another important previous reason- selectivity. A less potent inhibitor, which will have to be present at high concentrations to be effective, will be more likely to

exert off-target effects in the cell due to the great diversity and number of cellular proteins. Thus more potent compounds will usually also be more selective since they can be used at lower doses. A corollary of this fact is that a kinase inhibitor will have to demonstrate a minimum level of potency in order to be selective in the cellular milieu, and this will be irrespective of the *in vitro* data<sup>166</sup>. Both the above considerations place restrictions on the accurate design of kinase inhibitors.

### 5.6: The cell cycle, cyclin-dependent kinases and CDK7

The cell cycle is a series of events which all cells must traverse in order to divide (**Figure 5.3**). It is generally thought to be comprised of 5 states: G<sub>0</sub>, G<sub>1</sub>, S, G<sub>2</sub>, and M<sup>167</sup>.

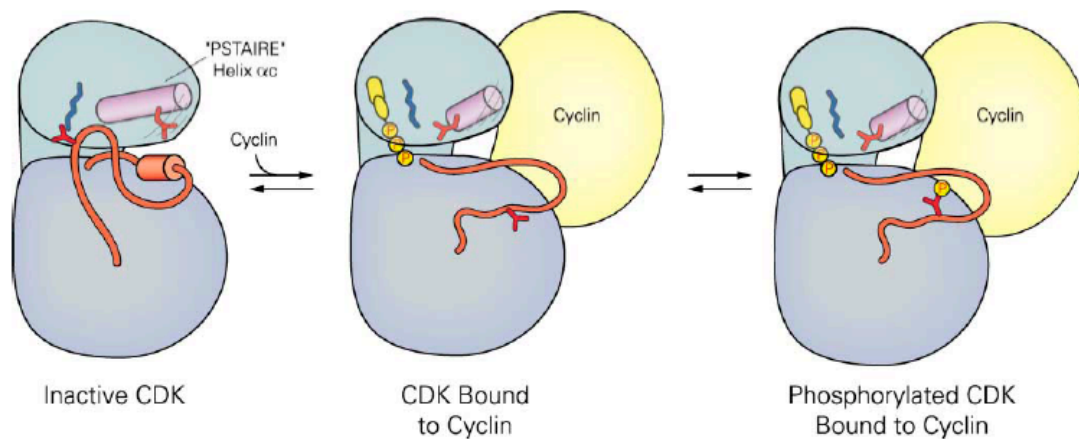


**Figure 5.3:** The cell cycle<sup>168</sup>

G0 is the quiescent state wherein cells are metabolically active but not embarking upon cell division. Upon stimulation with growth factors, cells leave G0 and enter the first gap phase called G1, during which they prepare to replicate their DNA. At a point late in G1, cells make an irreversible commitment to divide; this is called the restriction point (START in yeast). Beyond the restriction point, cells enter S-phase, where the genomic DNA is duplicated. Following DNA replication, cells enter a second gap phase called G2, in which they prepare to divide into two daughter cells. During this phase, cells survey the newly synthesized DNA for errors and ensure that DNA replication has proceeded to completion. Cell division, or mitosis, ultimately occurs during M-phase.

Control mechanisms in the cell cycle are called checkpoints. These checkpoints and the accompanying steps are tightly regulated and many important proteins are involved in maintaining the exact timing of these events. These proteins play roles in processes such as DNA repair and committing the cell to apoptosis if repair is not possible. Among these proteins, the cyclin-dependent kinases (CDKs) are key players that modulate the timing of the cell cycle through phosphorylation of substrate proteins and peptides<sup>169,170</sup>. These serine-threonine kinases play a special role in controlling the exquisitely timed actions of the cell cycle; they have been referred to as the ‘clocks’ and ‘microprocessors’ of the cell<sup>169</sup>. As their name indicates, the CDKs are dependent on the binding of proteins called cyclins for their activity. While the CDKs are present more or less throughout the duration of the cell cycle, cyclin levels are much more transient, thus carefully modulating the action of the CDKs. Central to the activation of a given CDK is the requirement for association with cyclins and phosphorylation at a threonine residue in the

activation loop (T-loop). Association with cyclins causes a rather large conformational change in the CDK in which a helix called the  $\alpha$ C helix swings a crucial glutamate residue towards a lysine in the active site (**Figure 5.4**)<sup>171</sup>. An ionic interaction between the glutamate and lysine catalyzes phosphate transfer and phosphorylation of substrate by ATP.



**Figure 5.4:** Activation of CDKs involves a conformational changes induced in helix  $C\alpha$  by cyclin binding (Figure from Huse et al.<sup>171</sup>)

Several CDKs annotated from CDK1 through CDK11 have been discovered and each one binds to a particular cyclin partner, with some binding to more than one. CDKs are important not just for their signal-transduction modulating ability but also because of their ability to control gene expression through phosphorylation.

In metazoans, phosphorylation of the CDKs that are required for cell cycle progression (CDK1, CDK2, CDK4, CDK6) is mediated by the CDK activating kinase (CAK), having three subunits CDK7, Cyclin H (CycH) and MAT1<sup>172-174</sup>.

Deregulation of CDK activity forms an important part of many cancers, as well as other disease states, generally through elevated and/or inappropriate activation, as CDKs are infrequently mutated<sup>167</sup>. Important mechanisms of CDK deregulation include cyclin over-expression, e.g. cyclin D1<sup>175</sup> and loss of CDKI expression through mutational or epigenetic alterations (for review see ref. <sup>176</sup>). As such, CDKs are important targets for the design of anti-cancer drugs. Some inhibitors of some CDKs, particular emphasis being placed on inhibitors of CDK2 as it controls S-phase entry, have been developed and some have been tested in the clinical setting as anti-cancer agents<sup>177-179</sup>. One of these is flavopiridol, which has modest selectivity for CDKs over other kinases and inhibits many members of the CDK family<sup>180</sup>. The compound class that has yielded many CDK-selective ATP antagonists is 2,6,9-trisubstituted purines, exemplified by roscovitine, which shows good biological and pharmacological properties<sup>181,182</sup>. CDK7 is an attractive target for drug development due to its critical role in the activation of the CDKs required for cell cycle progression<sup>172,174</sup>. This is especially significant as there is evidence that inhibition of some cell cycle CDKs may be compensated for by other CDKs. Hence, cells from mice that have been ablated for CDK2 are able to cycle, and CDK2<sup>-/-</sup> mice are viable<sup>183,184</sup>. Similarly, as are CDK4<sup>-/-</sup> and CDK6<sup>-/-</sup> mice are viable, although the double null mice show late embryonic lethality. However, mice lacking MAT1 die early in embryogenesis<sup>185</sup>, indicative of a cellular requirement for CAK.

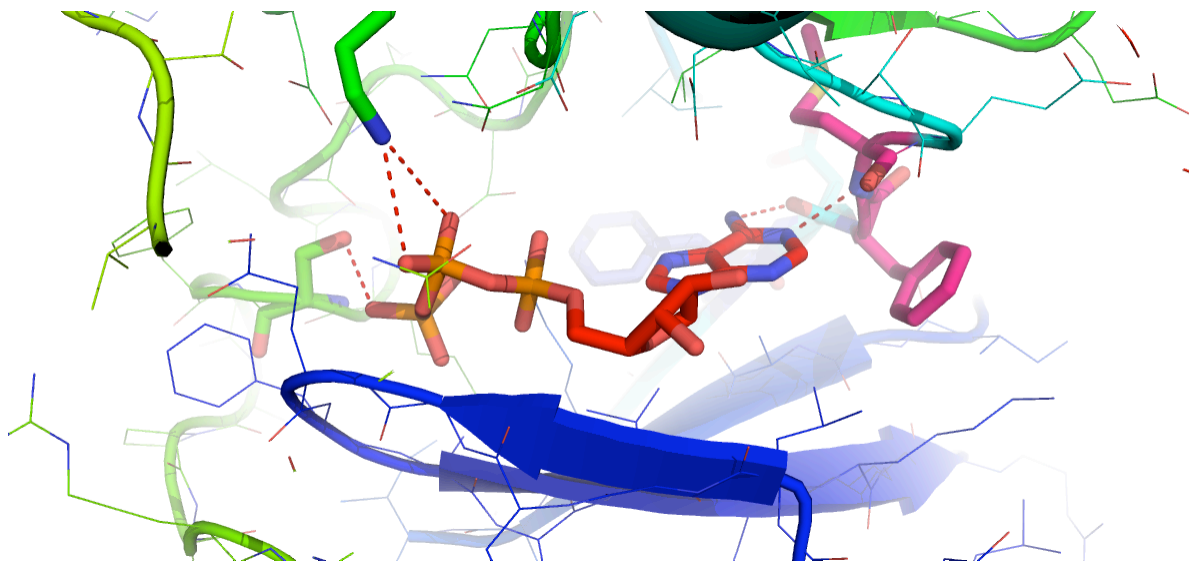
Although many of the CDK inhibitors that have been described inhibit CDK7, they only do so at higher concentrations<sup>182,186</sup> and some also inhibit other CDKs such as CDK5 and CDK9, which play important roles in neuronal development and transcription<sup>184,186-188</sup>. In addition to its role in cell cycle regulation, CDK7/CycH/MAT1 are components of the general transcription factor TFIIF<sup>189,190</sup>, required for initiation of transcription of RNA polymerase II (PolII)-directed genes. As part of the TFIIF complex, CDK7 phosphorylates the C-terminal domain of the largest subunit of RNA polymerase II<sup>172</sup>. Further, CAK or TFIIF-associated CAK phosphorylate several transcription factors to regulate their activities (e.g. see<sup>191,192</sup>). Inhibition of CDK7 activity would therefore be expected to inhibit transcription, as well as inhibiting cell cycle progression. There is however, evidence to suggest that inhibition of CDK7 activity is not sufficient to inhibit transcription. For example, inhibition of CDK7 activity in an engineered HeLa cell line did not reduce PolII phosphorylation at Ser-5, whereas phosphorylation of CDK1 and CDK2 was inhibited and cell cycle arrest was obtained<sup>193</sup>, whilst a kinase-dead mutant of CDK7 did not block TFIIF action in basal and activated transcription<sup>194</sup>. Moreover, there is no inhibition of transcription in MAT1<sup>-/-</sup> mice<sup>185,195</sup>, suggesting that CDK7 is indispensable for transcription, perhaps due to its function being compensated for by other kinases, such as CDK9<sup>196</sup>.

Selective inhibitors would provide potentially significant tools for dissecting further the multiple roles of CDK7 and could have utility as anti-cancer drugs.

### ***5.7: Design of selective CDK7 inhibitors***

Our interest in CDK7 was initiated by colleagues Simak Ali, Charles Coombes and Anthony Barrett from Imperial College London, UK. Earlier Coombes et al. had identified phosphorylation of the estrogen receptor as a key alternative mechanism activating the ER in tamoxifen-resistant breast cancer<sup>197</sup>. Tamoxifen is the world's leading drug against breast cancer and is in many cases the drug of choice<sup>198</sup>. It blocks estrogen binding to the ER and subsequent transcriptional activation. Yet up to 30% of patients develop resistance to tamoxifen through various mechanisms<sup>199,200</sup>. Some of these mechanisms involve activation of the ER through estrogen-independent processes<sup>197</sup>. CDK7 was identified as part of the TFIIF transcription factor machinery that phosphorylates Ser118 in the ER. In the study by Coombes et al., overexpression of CDK7 was found to stimulate transcriptional activation of the ER; this mechanism was identified as an alternative mechanism activating the ER and contributing to metastasis in breast cancer. Thus, molecules inhibiting CDK7 were thought to be potential useful drugs as well as tools for investigating details of CDK7 action.

The Imperial College scientists approached us with a view to potentially using computational methods for modeling the binding of inhibitors to CDK7 and possibly improving their binding profile. The starting point was roscovitine since roscovitine was already known to be a non-selective inhibitor of CDK7<sup>201,202</sup>. We decided to pursue a structure-based approach since the structure of CDK7 with bound ATP (**Figure 5.5**) and the structure of CDK5 with bound roscovitine had already been published<sup>203</sup>.



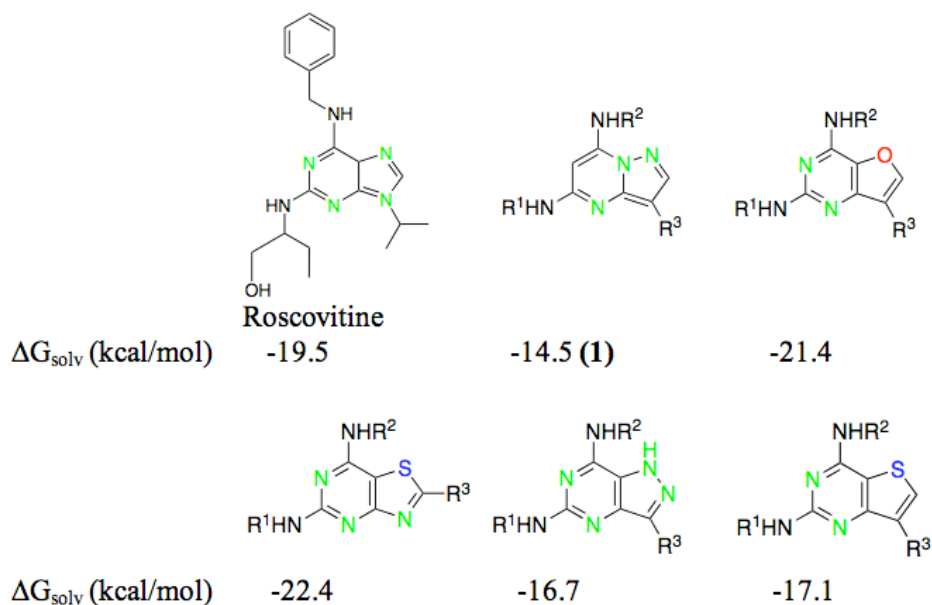
**Figure 5.5:** Crystal structure of CDK7 with bound ATP (PDB code: 1UA2). Hydrogen bonds with hinge residues Met and Phe and back residues Lys and Ser are shown as dotted red lines. Phe residue in stick form at the back of the adenine ring is gatekeeper. Hydrogen atoms have been omitted for clarity.

### **5.8: Computational modeling of BS-181**

Computational design began with selection of an alternative core heterocyclic ring structure able to preserve roscovitine side chain functionality, offer synthetic access and incorporate suitable solubility properties. Five motifs (**Figure 5.6**) were evaluated by the Amsol 6.6 package<sup>204</sup> with the expectation that the structure with the least favorable aqueous solvation energy would be transferred most readily into the hydrophobic kinase active site pocket. The calculated free energies of solvation suggested that pyrazolopyrimidine **1** would be soluble, but the most readily expelled from the water environment around the protein<sup>205</sup>. GLIDE (v 3.5) scores also indicated this motif to bind



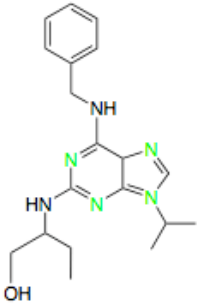
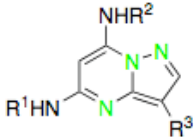
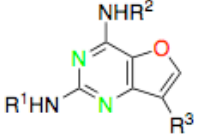
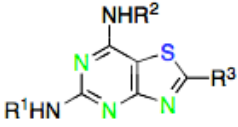
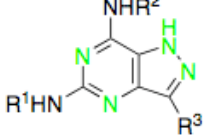
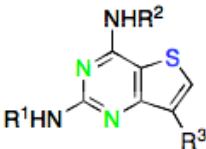
more favorably in the binding pocket of CDK7. Accordingly, this motif (**1**) was selected for synthetic modification.



**Figure 5.6:** AMSOL Solvation energies of six core motifs based on roscovitine.  $R^1$ ,  $R^2$  and  $R^3$  are identical to those in roscovitine

The AMSOL solvation energies for the motifs along with the GLIDE scores are listed in

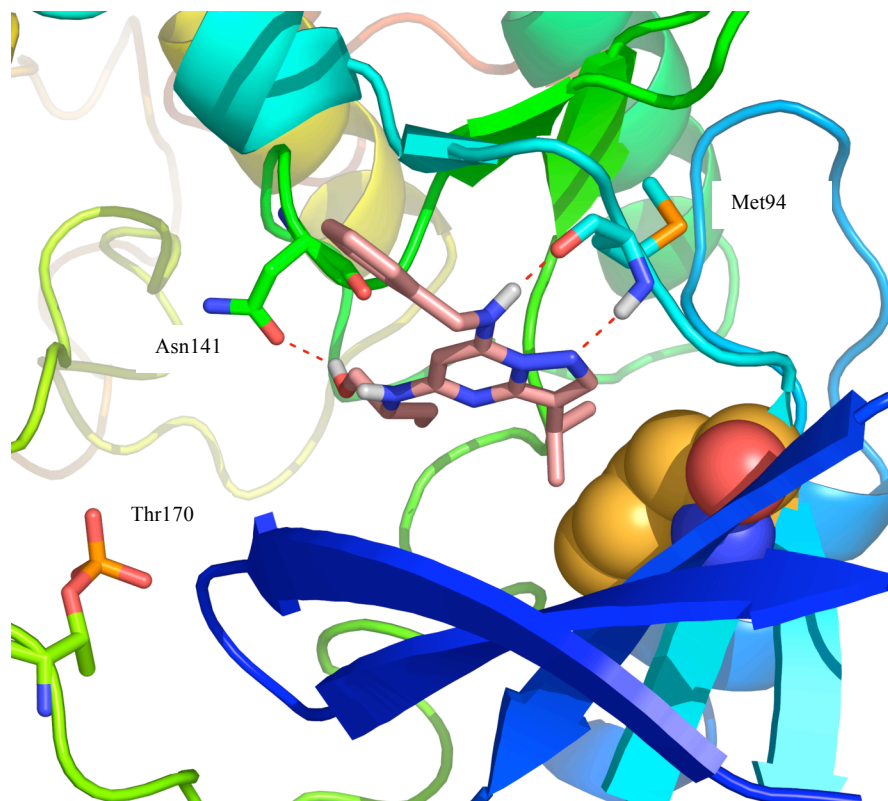
**Table 5.2**

Compound	Glide XP docking score (kcal/mol)	Solvation energy (kcal/mol)
 <p>Roscovitine</p>	-7.7	-19.5
	-9.3	-14.5
	-7.4	-21.4
	-5.5	-22.4
	-7.9	-16.7
	-6.3	-17.1

**Table 5.2:** Docking scores and solvation energies of core templates

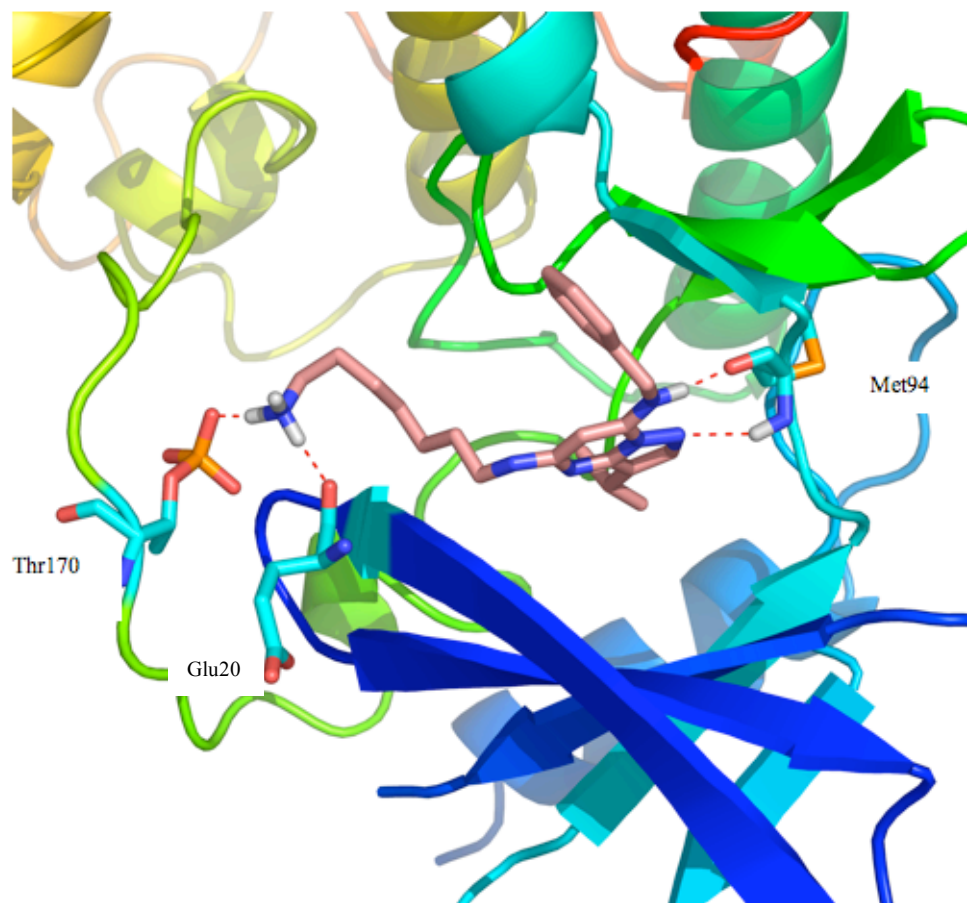
As a benchmarking exercise, roscovitine itself was removed from the CDK5 active site (PDB ID: 1UNL, resolution 2.20 Å) and redocked using the Glide SP<sup>206</sup> module from Schrödinger (v 7.5). Excellent overlap of the resulting structure with the original crystallographic pose (RMSD 0.1 Å) indicated Glide docking to be an effective tool for further structure-based design in the present context. This ‘self-docking’ procedure has previously been used to gain confidence in the validity of docking programs for particular systems<sup>206</sup>. When the six templates complemented with roscovitine substituents (Figure 2) were docked into the CDK7 kinase pocket (PDB ID: 1UA2, resolution 3.0 Å) using the same protocol, **1** gave the best Glide scores compared with the other templates (Table 1). Past studies have indicated that while absolute values of scores from scoring functions are not reliable, relative values especially for congeneric series with a known reference can provide insight into ranking ligands by binding affinity<sup>39,207</sup>. Thus the choice of **1** was dictated not only by solvation energies but also by docking scores. In subsequent docking exercises we employed Glide XP, a Glide variation designed to more accurately capture the basic physical chemistry of ligand binding. The method has been well documented as a tool for lead optimization<sup>206,208</sup>. All docking runs were performed with default settings including settings for protein preparation (assignment of protonation states and addition of hydrogens), generation of a grid centered on the ligand, and docking in SP and XP modes.

A key observation in preliminary studies was that **1** with the same side chains as roscovitine docked into the CDK7 active site in the same orientation as roscovitine but with slightly better scores and substantially more favorable solvation energies. The best pose for **1** is similar to that for other pyrazolopyrimidines.<sup>6</sup> The 3-isopropyl group protrudes into the cavity formed by the gatekeeper Phe91, N1 and N6 form hydrogen bonds with backbone atoms of Met94 in the hinge region of the kinase and the side chain hydroxyl makes a hydrogen bond with Asn141 (**Figure 5.7**)



**Figure 5.7:** Lowest energy GlideXP pose for pyrazolopyrimidine **1**. The orange space-filling Phe91 is the gatekeeper residue. Hydrogen bonds with Met94 and Asn141 are displayed as dashed lines.

The docked pose suggested that **1** is incapable of completely occupying the CDK7 active site pocket. Left unutilized was a sector in the back of the cleft occupied by two lysines (Lys41 and Lys138) and a phosphorylated threonine (Thr170). In an attempt to exploit the lipophilic nature of this subsite, the hydroxy ethyl moiety of **1** borrowed from roscovitine was excised; and the resulting propyl side chain, extended. Nonpolar alkyl linkers of different chain length terminating in a variety of polar groups attached to the NH at the C2 position were conceived and docked. A six-carbon linker with a primary amine terminus (Figure 3) delivered the best docking score and suggested a considerable affinity improvement relative to roscovitine. At the same time, the corresponding Glide pose was similar to that for **1**. In addition to hydrogen bonds in the hinge region, the protonated distal amine was predicted to participate in strong electrostatic interactions with the phosphate group of Thr170 and the backbone carbonyl of Glu20. Simultaneously, the six-carbon linker exhibited productive van der Waals contacts with the floor of the kinase binding pocket (**Figure 5.8**). The same structure was estimated to possess a favorable solvation energy.



**Figure 5.8:** Glide XP pose for BS181 with hydrogen bonds to Met94, Thr170 and Glu20.

Additional docking studies also indicated the structure to favor CDK7 relative to CDK2 (PDB ID: 1B38), CDK5 (PDB ID: 1UNL) and CDK6 (PDB ID: 1XO2) suggesting possible selectivity (**Table 5.3**)

Protein	Glide XP score (kcal/mol)
CDK 7	-13
CDK 2	-9
CDK 6	-9
CDK 5	-11

**Table 5.3:** Comparison of Glide XP docking scores for BS181 in CDK7, 2, 5 and 6.

The relative docking scores were confirmed and replicated by induced-fit docking using the Prime module from Schrödinger (default settings). Encouraged by these results, the structure now named BS-181 was subjected to synthesis and bioassay.

The modeling predictions were substantiated by experiment when BS181 turned out to be a more potent and selective CDK7 inhibitor than roscovitine. For example, as shown here, roscovitine is an inhibitor of cyclin dependent kinases 2, 5, 7 and 9, with  $IC_{50}$  values of 100, 240, 510 and 1200 nM, respectively. BS-181, on the other hand, exhibits a substantially higher preference for CDK7 with an  $IC_{50}$  value of 21 nM. Excellent selectivity against CDK2, CDK5 and CDK9 is illustrated by high  $IC_{50}$  values of 880, 3000 and 4200 nM, respectively (**Table 5.4**). BS-181 also fails to block CDK1, 4 and 6, with  $IC_{50}$  values being greater than 5000 nM. As such, BS-181 is a highly selective CDK inhibitor, and is the most potent CDK7 inhibitor described to date.



Kinase	Roscovitine		BS-181	
	IC <sub>50</sub>	(SD) (μM)	IC <sub>50</sub>	(SD) (μM)
CDK7	0.51	(0.1)	0.021	(0.002)
CDK1	1.8	(0.3)	8.1	(0.6)
CDK2	0.1	(0.02)	0.88	(0.08)
CDK4	15.3	(6.6)	33	(1.5)
CDK5	0.24	(0.1)	3.0	(0.5)
CDK6	28	(4.9)	47	(4.0)
CDK9	1.2	(0.8)	4.2	(0.5)

**Table 5.4:** BS-181 is more selective for CDK7 compared to roscovitine (data from Ali et al.<sup>209</sup>)

### 5.9: Biological studies with BS181

The biological studies were performed by the Imperial College biologists and the figures, captions and material in this section are as described by these researchers in reference 203.

BS-181 was selective for CDK7 over several other CDKs (**Table 5.4**). In addition, in cell-based assays with cell lines from different tumors, it showed comparable activity to roscovitine (**Table 5.5**)

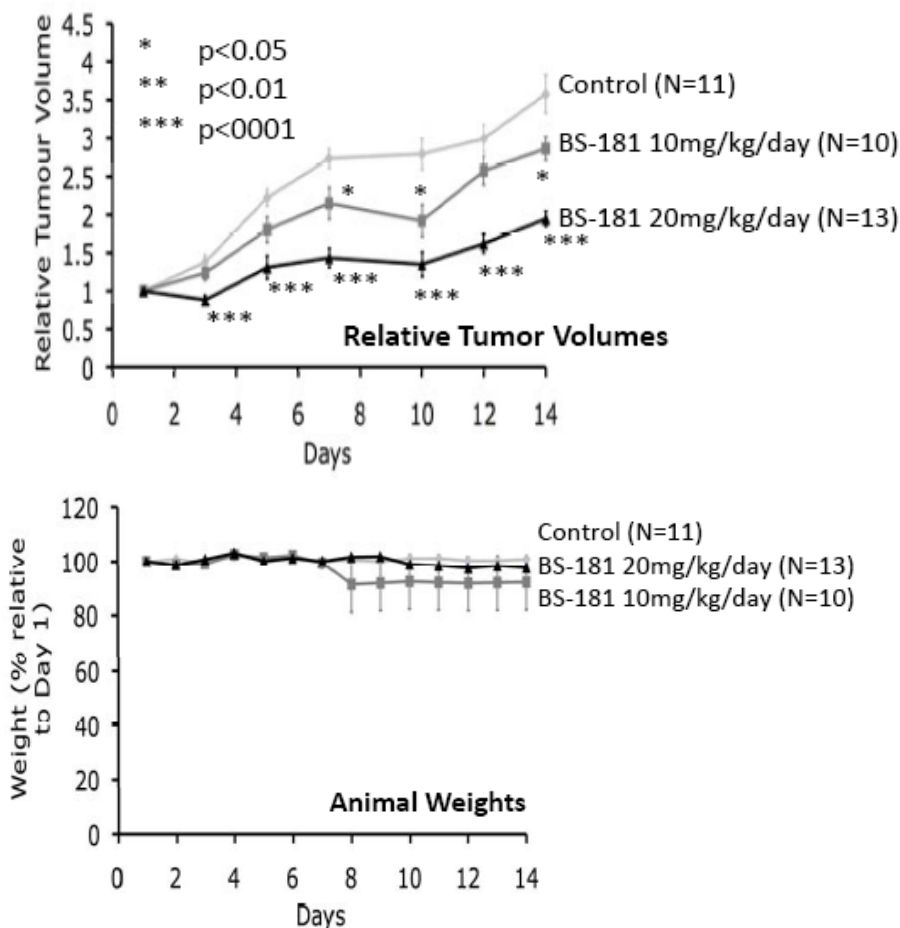
Tumour Type	Cell line	Roscovitine IC <sub>50</sub> (μM)	BS-181 IC <sub>50</sub> (μM)
Breast	MCF-7	13	20
	MDA-MB-231	18	15
	T47D	25	16.5
	ZR-75-1	33.5	25.5
	BT474	30.5	30.5
	BT20	32.5	19
Colorectal	COLO-205	8	11.5
Lung	A549	17	37
	NCI-460	9.5	21
Osteosarcoma	U2OS	10.5	14.5
	SaSO2	13.9	20
Prostate	PC3	10.8	23
	LNCaP	8.4	32
Liver	HepG2	12.3	24

The IC<sub>50</sub> values (μM) were obtained using the sulforhodamine B assay. Shown are the mean values derived from at least three replicates.

**Table 5.5:** GI<sub>50</sub> values for BS181 compared to roscovitine in various cell lines.

Intraperitoneal injection of BS-181 inhibited MCF-7 tumor xenografts, lending further support to a potential utility of CDK7 inhibitors in cancer treatment (**Figure 5.9**).

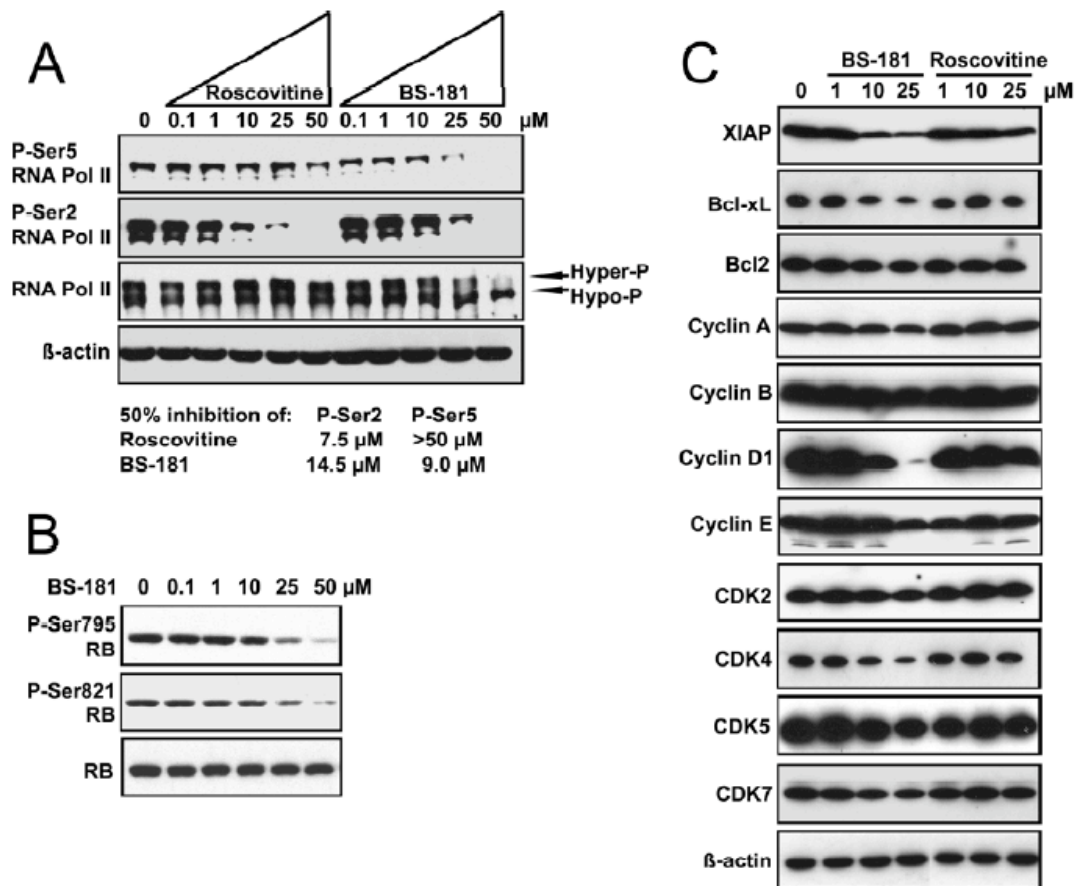
Pharmacokinetic studies showed rapid clearance of BS-181 administered i.p. or i.v. In the case of intraperitoneal administration, bioavailability was only 37%, indicating a need for further refinement of the BS-181 structure to improve stability and bioavailability. As it stands the studies described here indicate that continuous i.v. infusion or repeated administration is needed for further in vivo evaluation. The observed efficacy, despite the low plasma levels (lower than the  $IC_{50}$  for growth inhibition in vitro), could therefore be due, at least in part to more active metabolites generated following intraperitoneal administration.



**Figure 5.9:** BS-181 inhibits the growth of MCF-7 tumors in nude mice. Randomized MCF-7 tumor bearing mice were injected intraperitoneally twice daily with 5mg/kg or 10 mg/kg BS-181, giving a total daily dose of 10 mg/kg/day or 20 mg/kg/day, respectively, over a period of 14 days. Mouse weights were determined daily, tumor volumes being measured every 2 days (top). The change in tumor volume was determined for each animal, as tumor volume relative to the tumor volume of each animal at day 1. The line graphs show the mean tumor volumes for the animals in each treatment group. Error bars represent the standard error of the mean. Asterisks depict the statistical significance of the differences between the control group and each of the BS-181 treatment groups (bottom). Shown are the animal weights, as percentage change relative to the animal weights at day 1. Error bars represent the standard error of the mean.

To confirm that BS-181 retards the growth of tumors through inhibition of CDK7, inhibition of phosphorylation of the amino acid residue Ser5 on RNA polymerase II, a known CDK substrate, was monitored by radioassays. As shown in **Figure 5.10**, phosphorylation of Ser5 was inhibited at physiologically relevant concentrations. However, although CDK7 does not target Ser2, BS-181 also inhibited Ser2 phosphorylation, likely through inhibition of CDK2 and CDK9. Phosphorylation retinoblastoma protein, which is a substrate for CDK2, was also observed. At the same time, Ser5 inhibition was observed at lower concentrations of BS-181 than the concentrations required for inhibition of Ser2 phosphorylation. In this respect, roscovitine, which is a more potent inhibitor of CDK2 and CDK9 than BS-181, inhibited Ser2 phosphorylation at lower concentrations than BS-181, but only poorly inhibited Ser5

phosphorylation. Together, these findings suggest that while there is some inhibition of CDK2 and CDK9, CDK7 is a key target of BS-181 in MCF-7 cells.



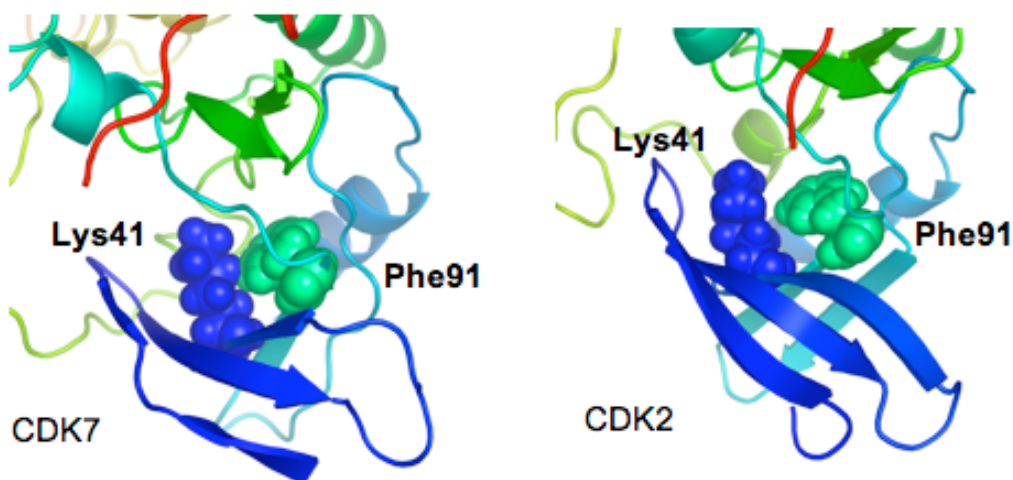
**Figure 5.10:** BS-181 inhibits phosphorylation of CDK7 substrates. Whole cell lysates were prepared from MCF-7 cells treated with BS-181 or Roscovitine for four hours, at the concentrations shown. (A) Immunoblotting was carried out using antibodies for RNA polymerase II, or Pol II phosphorylated at Ser2 or Ser5 in the C-terminal domain. The concentration at which apparent inhibition of PolII phosphorylation by 50% would be achieved was determined following densitometry of immunoblots from three experiments. (B-C) Immunoblotting was carried out as in part A, using the antibodies as labeled.

### ***5.10: Rationalization of CDK7 selectivity and implications***

It is difficult to rationalize computationally the selectivity of BS-181 for CDK7 over CDK2 and CDK5 in terms of specific ligand-protein interactions. In the past, several factors responsible for kinase selectivity have been invoked in specific cases. These include loop mobility and dynamics<sup>210</sup>, mutations in key residues and especially the gatekeeper residue<sup>211</sup> and variable solvation for different inhibitors<sup>209</sup>. In this particular case, reasons for selectivity are not obvious. However, clues from previous studies regarding the roles of water molecules in hydrophobic protein active sites have provoked and enabled us to construct a hypothesis for explaining selectivity. It has been well-documented that water molecules in tightly packed hydrophobic environments can have both enthalpically and entropically unfavorable characteristics; typically such waters have unfavorable entropy due to the constraining nature of the packed hydrophobic environment and unfavorable enthalpy because of their inability to form a full complement of four hydrogen bonds in the absence of surrounding polar groups<sup>205,212</sup>. Displacement of such waters by hydrophobic ligand functionalities can contribute substantial gains to binding affinity because of the resulting entropic and enthalpic freedom<sup>205,213,214</sup>. In this specific case, different packing interactions of the non-polar isopropyl side chain at C3 in BS-181 with the amino acids in the kinase pocket may aid in explaining the phenomenon. For example, the 3-isopropyl side chain protrudes into a cavity formed in part by the important gatekeeper residues Phe91 and the C4 carbon chain of Lys 41 in both CDK2 and CDK7. However, the hydrophobic packing of the two residues is much tighter in the case of CDK7 than it is in CDK2 (**Figure 5.11**). This volume-based realignment in the gatekeeper sector of the binding site may well exert a

subtle effect that influences selectivity. To test this hypothesis, we are engaging in a possible collaboration with Schrödinger in order to utilize their WaterMap<sup>213</sup> application. WaterMap performs a molecular dynamics simulation with the program Desmond<sup>215</sup> and maps the thermodynamic characteristics of water molecules in active sites by dissecting their free energy into enthalpic and entropic terms. Thus one can get a precise picture of the differences in thermodynamic variables of water molecules in the two active sites. The presence of more thermodynamically unfavorable water molecules in CDK7 compared to CDK2 will provide support for the selectivity hypothesis.

The hypothesis also sheds light on the interactions of the gatekeeper residue with the C3 side chain. If the packing hypothesis is indeed correct, larger substituents might provide even more selectivity for CDK7 over CDK2 by encouraging better packing with the hydrophobic residues.

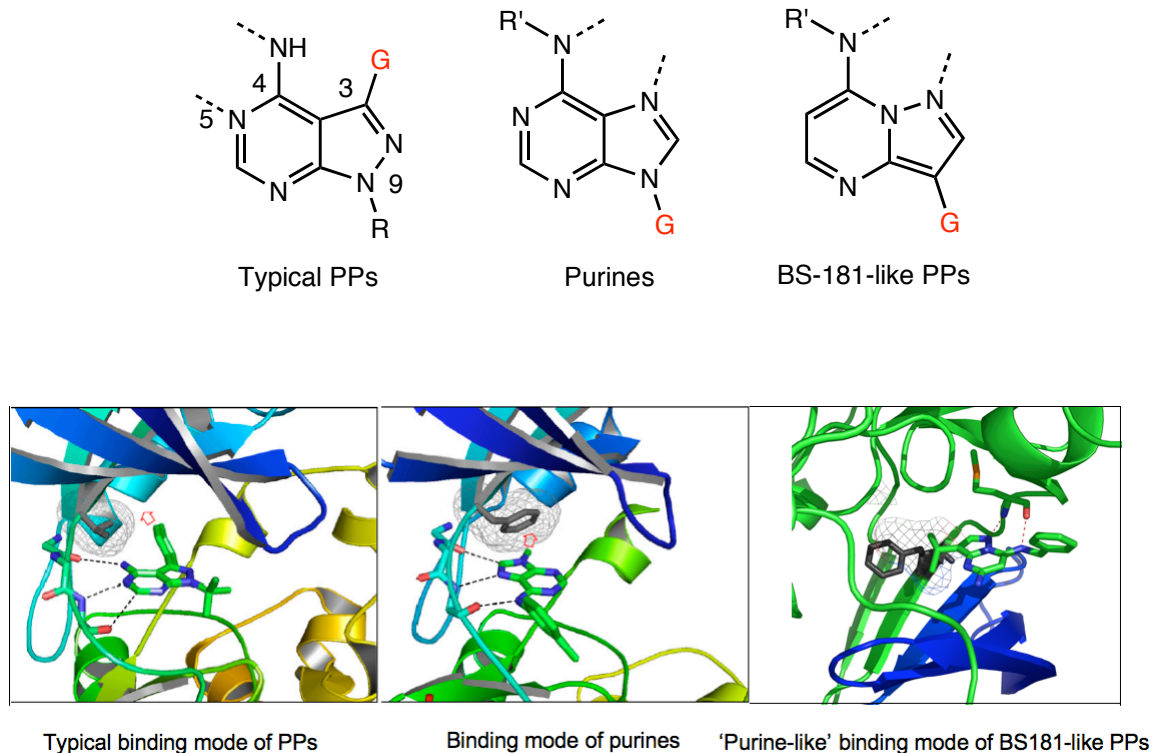


**Figure 5.11:** Packing interactions between gatekeeper residue Phe91 and Lys41 in CDK7 (left) and CDK2 (right)



The differential packing interactions also highlighted a rather unique property of BS-181 and related compounds that may aid future design for enhancing selectivity and potency. Pyrazolopyrimidines (PPs) such as BS-181 are known inhibitors of protein kinases such as p38 MAP kinase and some have successfully entered clinical trials. However, most compounds defined as PPs have a structure different from BS-181 and related compounds, a chance observation made during the perusal of a paper by Zhang et al.<sup>211</sup>

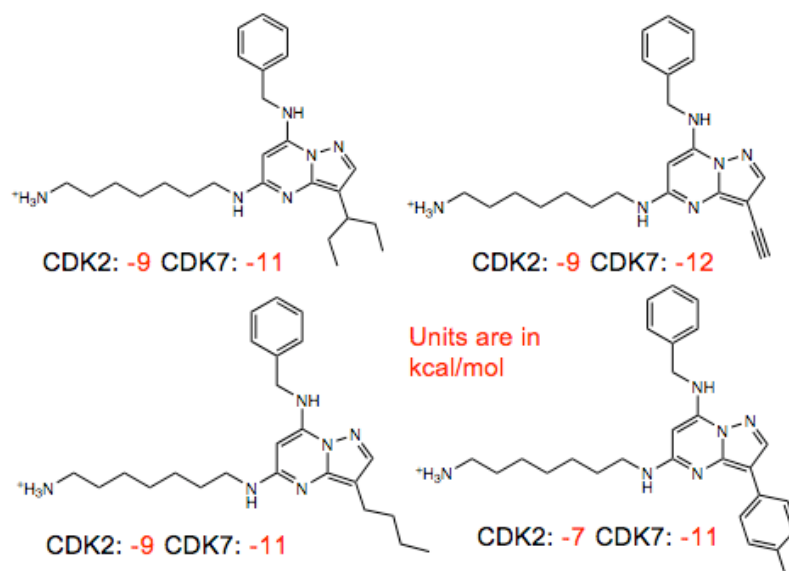
Most PPs have a nitrogen at the 5 position unlike BS-181 and do not have a nitrogen at the 3 position. However, purines have a nitrogen at both the 3 and 5 position. Either one of these nitrogens along with the 4-amino NH can interact with the hinge region residues. The choice of interacting nitrogen has a profound effect on the binding mode with respect to proximity to the gatekeeper residue. In case of purines, N3 and the 4-NH interact with the hinge residue, which locates the N9 substituent close to the gatekeeper. In case of typical PPs it's usually N5 and the 4-NH which interact with the hinge region and thus locate not the N9 substituent but the C3 substituent close to the gatekeeper. The significant difference between typical PPs and BS-181-like compounds is the absence of N5 and presence of N3 in the latter. This subtle difference has a profound effect on the binding mode of BS-181; in the absence of N5 only N3 can interact with the hinge, thus rendering the binding mode of BS-181 'purine-like' (**Figure 5.12a and 5.12b**).



**Figure 5.12: a.** Differences between binding site interactions of typical PPs, purines and BS-181-like PPs. The red substituent corresponds to the substituent interacting with the gatekeeper and the dotted lines indicate the atoms interacting with the hinge region. **b.** The same differences illustrated by interactions in the binding site (Figure adapted from Zhang et al.<sup>211</sup>)

In this orientation it's the C9 isopropyl substituent that protrudes into the gatekeeper residues. Modification of this substituent combined with the knowledge of the difference in packing interactions in this region noted above should provide an opportunity for enhancing selectivity. Based on this hypothesis, there are ongoing synthetic investigations to replace the isopropyl side chain with bulkier groups like *s*-butyl and

isobutyl, which may render better selectivity for CDK7. Preliminary docking runs indicate the possibly viability of this hypothesis for a CDK7/CDK2 comparison (**Figure 5.13**)



**Figure 5.13:** GLIDE (v 3.5) docking scores for BS-181 analogs with different substituents at the C9 position.

### ***5.11: Outlook and future directions***

During this study, which began in 2005, we have discovered the most potent and selective inhibitor of CDK7 known to date, BS-181. While the compound potently inhibits CDK7, it does not possess optimum pharmacokinetic properties and especially suffers from poor bioavailability and cell permeability. Strategies to improve these parameters are underway and include modification of the phenyl ring and investigations into prodrug design. Studies to evaluate the effects of metabolites are also ongoing. The phenyl ring can purportedly be safely modified with small substituents since it projects into solvent and does not form crucial interactions with amino acids residues which could be detrimentally perturbed by addition of substituents. Decorating the ring with polar substituents like fluorine and sulfonamide may also improve stability and solubility. However, there is also a parallel effort that has been initiated to investigate BS-178, another roscovitine modification which was discovered early to be a pan-CDK inhibitor. While this compound is not selective for CDK7, it is about a hundredfold selective for CDK5 over other CDKs, but more importantly has improved pharmacokinetic properties over BS-181. Since many clinical candidates fail due to poor ADMET (Absorption Distribution Metabolism Excretion Toxicity) properties<sup>216</sup>, it is worthwhile advancing studies with candidates which may not be especially potent but which may be potentially better tolerated in patients. To this end, it is prudent to continue studies both with the more potent BS-181 as well as the less potent and more promiscuous BS-178 which nonetheless seems to have a better PK profile. Also as noted in the discussion above, lack of potency of a compound does not necessarily translate into a poor prognosis for clinical studies. As modifications of BS-181 and studies with BS-178 continue, only time will tell

if one of these candidates demonstrates all the properties necessary for a drug molecule to enter the clinic, and traverses all the pitfalls and challenges associated with the long and winding road to the market.

# Chapter 6

## Self-assembly of amyloid A $\beta$ (16-22)

“ Make everything as simple as possible, but not simpler”

-Albert Einstein

### **6.1: Alzheimer's disease**

**A**mong all the disorders and diseases afflicting humanity, neurodegenerative diseases pose a special burden. The relatively slow but certain onset of these diseases, the unusual emotional pain associated with their manifestation and treatment and the lack of cures for these diseases make them among the most devastating diseases to affect society, in terms of both economic and human costs. Foremost in purview and magnitude among these diseases is Alzheimer's disease (AD)<sup>217,218</sup>. From the time that AD was first observed and noted in patient Auguste D. by German neurologist Alois Alzheimer in 1906<sup>219</sup>, the disorder has come to occupy significant resources of scientists, doctors, healthcare officials and families around the world.

AD is characterized by cognitive impairment and memory loss<sup>218,220</sup>. Early symptoms include confusion and forgetfulness and a gradually increasing inability to perform simple daily tasks. Surprisingly, long-term memory loss is not seen during early stages of the disease. As the disease progresses, there is a pronounced loss of motor skills and increased long-term memory loss. In the late stages of AD the patient becomes almost completely senile and unable to perform the simplest functions necessary for living a normal life. At this point there is complete dependency on a caregiver<sup>221</sup>. Death usually results not from the disease itself but from accompanying complications such as pneumonia and breathing difficulties<sup>222</sup>. As an added source of personal trauma, AD usually takes several years to manifest itself, a period during which its effects are gradually and incrementally amplified.

Since AD primarily affects the elderly, its toll has increasingly spread in the last several decades, even as nutrition and better medication have extended the average lifespan of many around the globe. Once an affliction of mainly developed countries, AD is now a growing concern in developing nations<sup>223</sup>. Part of the reason AD is widely feared is its certain onslaught and progress; while a cure for AD seems impossibly far on the horizon, even treatments designed to slow the its progress seem disappointing and have failed to live up to their promise<sup>224</sup>. Another cause for the grim prognosis of AD is the lack of a test for early diagnosis because of which detection and treatment both come too late for any possible significant results<sup>224</sup>.

Late detection and a complete lack of a treatment or cure make investigations of the causes, progression and treatment of AD one of the most urgent priorities for biomedical research. The urgency is exemplified by the number of AD cases around the world; about 26.6 million cases worldwide in 2006, with this number expected to quadruple by 2050<sup>225</sup>. Because of the dependence engendered by AD, patient management in AD is crucial and this is expected to exact a very large burden on caregivers; the burden can consist of significant economic, personal, physical and psychological sacrifices and contributions on the caregiver's part<sup>221</sup>. In developed countries AD has become one of the most economically costly diseases.

Many studies over the last few decades have sought to unravel the causes and risk factors associated with AD. Heredity is now definitely thought to play a role<sup>217</sup>. For instance the existence of the APOE4 genetic allele that codes for the E4 isoform of apolipoprotein-E

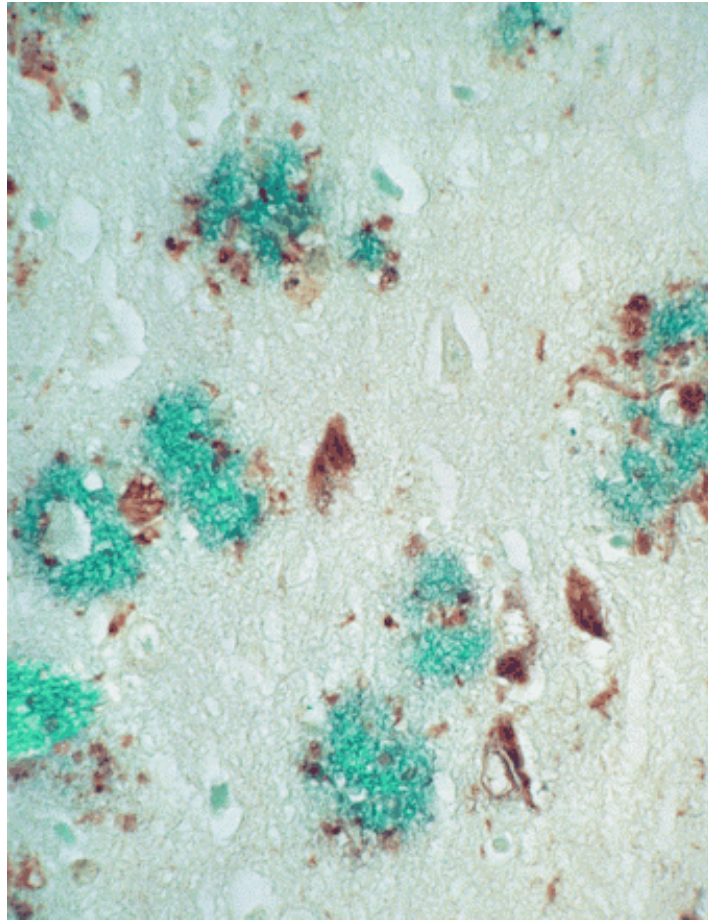


has been found to be strongly correlated to the risk of getting AD. A strong basis for this hypothesis comes from studies of people with Down's syndrome (trisomy 21)<sup>226</sup>. The APOE4 allele is located on chromosome 21. Down's syndrome patients who have two copies of the allele are almost universally known to be at increased risk of getting AD by age 40. Caucasians and Japanese populations who have two copies of this allele are thought to be 10 to 30 times more susceptible to getting afflicted with AD by age 75 and onwards compared to those who lack a copy of APOE4. The exact relation of the APOE4 allele and AD is not known, although it is thought to involve defective processing of the key amyloid precursor protein (APP) that apolipoprotein-E usually engages in.

Apart from genetics, dietary factors and metals that cause oxidative damage are also thought to be connected to a higher risk of AD<sup>227</sup>. Another hypothesis is based on decreased synthesis of acetylcholine in neurons; this hypothesis is now under suspicion, however, since drugs that potentially target acetylcholine deficiency have failed to demonstrate efficacy<sup>224</sup>. The NMDA and metabotropic glutamate receptors have also been implicated in AD. However, the search for non-genetic causes of AD is still in its early stages and much progress remains to be made.

## ***6.2: Amyloid A $\beta$ and Alzheimer's disease***

It is however the study of anomalous protein deposits in AD that has proven most compelling from a fundamental scientific and causal viewpoint. Early on when Alois Alzheimer documented AD in Auguste D., he observed insoluble tangled plaques of protein in her brain during the autopsy<sup>219</sup> (**Figure 6.1**).

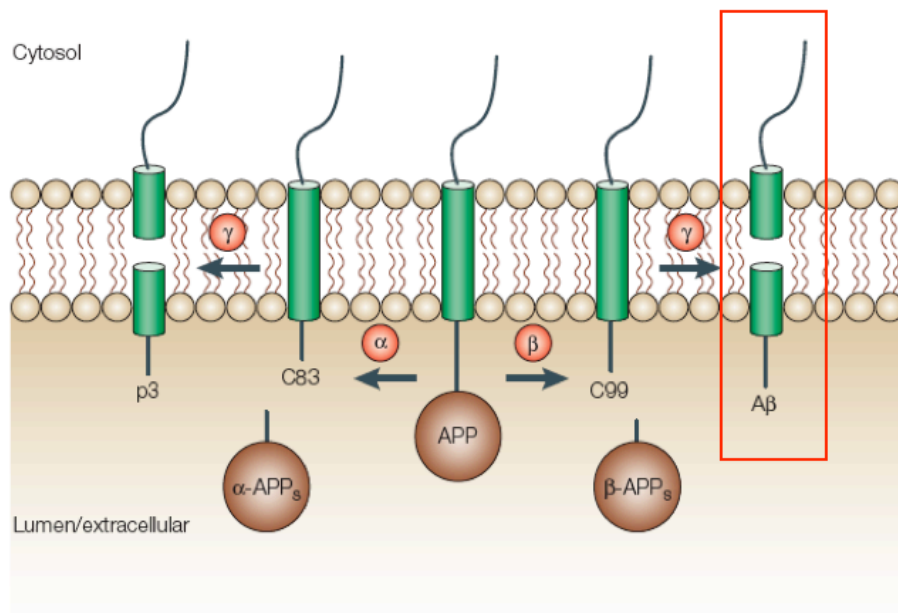


**Figure 6.1:** Histopathological image of stained amyloid plaques (blue) and tau neurofibrillary tangles (purple) in the brain of a person with hereditary AD (Figure from Kumar-Singh et al.<sup>228</sup>)

This protein was observed and studied over the next few decades. After sequencing and investigating it at the molecular level, the protein was found to consist of 40 or 42 amino acid residues and was named amyloid A $\beta$  [A $\beta$  (1-40) or A $\beta$  (1-42)]. Later studies indicated the ubiquitous presence of this protein aggregate in the brains of AD patients and around 1991, a hypothesis called the ‘amyloid cascade hypothesis’ emerged to

suggest an association between amyloid A $\beta$  and AD<sup>229</sup>. While another important hypothesis based on observation of neurofibrillary tangles of a protein called tau has also been suggested<sup>230</sup>, it is the amyloid cascade hypothesis that has constituted the frontline paradigm in understanding AD at the molecular level.

From a pathological perspective, AD is characterized by neuronal and synaptic death in key areas of the brain that play important roles in memory and cognition, such as the cerebral cortex<sup>231</sup>. According to the amyloid cascade hypothesis, the presence of toxic amyloid aggregates in the brain is strongly associated with the neuronal death and cognitive impairment observed in AD. Detailed molecular studies have revealed that A $\beta$  results from the processing of the amyloid precursor protein (APP), a membrane protein whose function is still unknown. APP undergoes proteolytic cleavage by three proteases; termed  $\alpha$ ,  $\beta$  and  $\gamma$  secretases. All three secretases cleave APP in its extracellular domain and result in three kinds of diverse peptide fragments. Cleavage by  $\alpha$ -secretase results in soluble peptides that do not seem to possess toxicity. However, cleavage of APP by  $\beta$ -secretase followed by  $\gamma$ -secretase processing results in the toxic A $\beta$  (1-40) and A $\beta$  (1-42) fragments whose presence is highly correlated with AD (**Figure 6.2**).



**Figure 6.2:** Proteolytic cleavage of the membrane protein APP by  $\alpha$ -secretase produces fragment C83. Proteolysis by  $\beta$  secretase produces fragment C99 which undergoes cleavage by  $\gamma$ -secretase resulting in formation of A $\beta$  (Figure adapted from Wolfe<sup>232</sup>)

In the last two decades, A $\beta$  and its precise role in AD have received intense scrutiny. These studies have resulted in a picture of the association between amyloid and AD that is still dynamic and evolving. For a long time the principal question has centered on whether AD is caused by or correlated to the presence of amyloid, and what role if any the protein aggregates play in progression of the disease. Initial pathological and other studies indicated that fully aggregated insoluble A $\beta$  might be the main toxic species or even causative agent in AD. However, recent research spanning the last decade have emphatically sought to revise this hypothesis. Antibody labeling and related techniques have made the isolation of soluble oligomeric species of A $\beta$  possible and in recent years,

many studies have pointed to these species and not the insoluble aggregates as the true toxic agent in AD<sup>233,234</sup>. The exact role of insoluble A $\beta$  and soluble oligomeric A $\beta$  are still debated but some studies have conjectured that fully aggregated A $\beta$  might act as a reservoir of soluble oligomeric toxic species<sup>235</sup>.

Based on the current molecular level understanding of AD, there has been a substantial effort both in the public and the private research sector to target specific molecular events in AD. Thus, therapies have been developed to target the cholinergic pathway, NMDA receptor, enzymes involved in tau hyperphosphorylation<sup>236</sup>,  $\beta$  and  $\gamma$  secretase<sup>237</sup>, and most importantly, A $\beta$  assemblies. These therapies have included both small molecules as well as targeted antibodies<sup>232,238</sup>. Small molecule disrupters of A $\beta$  have garnered significant recent attention<sup>239,240</sup>. In spite of these efforts, the efficacy of targeting these molecular pathways is not clear as exemplified by the failure of two recent large-scale clinical trials<sup>224</sup>. One involved an inhibitor of  $\gamma$  secretase<sup>241</sup> while the other involved a peptide vaccine that would enable the body to recognize and dispose of misfolded A $\beta$ . Incidentally in the second trial, A $\beta$  burden did seem to be reduced, but without any improvement or stasis in cognitive capabilities. Consequently there is renewed doubt about whether A $\beta$  is an important part of the etiology of AD. Thus the present consensus seems to be that a reexamination of the relationship between these molecular events and the symptoms of AD is necessary and focus needs to be diverted toward other relatively neglected avenues, including tau formation.

While therapeutic efforts toward AD continue, so do basic biostructural investigations of amyloid which are valuable in themselves for providing fundamental insight into protein structure and function. One of the principal challenges in studying amyloid is its aggregated nature because of which it is extremely hard to obtain well-resolved regular crystals of the protein. Consequently, x-ray diffraction by itself is not adequate to reveal detailed structural information at the atomic level, although it can provide valuable coarse information. Nonetheless, studies by techniques like solid-state NMR, small and wide angle x-ray diffraction, fluorescence and electron microscopy have revealed amyloid A $\beta$  to possess an archetypal laminated  $\beta$ -sheet structure called the ‘cross  $\beta$  structure’<sup>242</sup>. This structure reveals itself through characteristic x-ray reflections at about 5 Å and 10 Å which correspond to the repeat hydrogen bonding distances between two beta sheet strands and the lamination distance between two beta sheets respectively<sup>243,244</sup>. One of the most interesting discoveries of the last two decades concerns the observation of amyloid formation in many other proteins other than A $\beta$ , with each one being defined by the characteristic cross  $\beta$  structure and corresponding x-ray reflections<sup>245</sup>. Some proteins can even transform into amyloid from their folded state under native conditions<sup>246</sup>. Thus among others, insulin (Type 2 diabetes), prion protein PrP (transmissible spongiform encephalopathy or ‘Mad Cow Disease’),  $\alpha$  synuclein (Parkinson’s disease) and apolipoprotein AI (Artherosclerosis) are all known to form amyloid under the right physicochemical conditions<sup>247,248</sup>. This fact points to the cross  $\beta$  structure being a minimum on the folding energy surface of the protein. Far from being an anomaly, amyloid seems to be a common conformational motif accessible to diverse and common

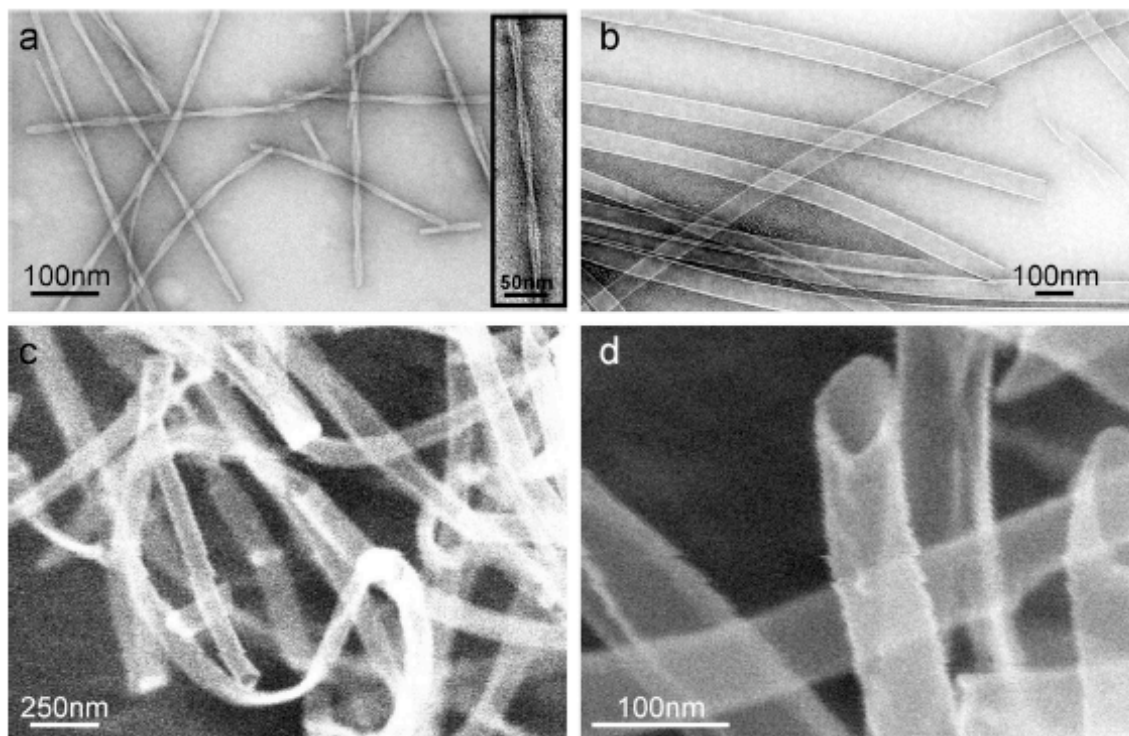
proteins. Interest has thus centered on the elucidation of the factors responsible for formation of amyloid among various proteins.

### ***6.3: Self-assembly of A $\beta$ (16-22)***

Our interest in amyloid was generated by the pioneering investigations of David Lynn's group at Emory University. The Lynn group has devoted considerable attention to elucidating the physicochemical reasons prevalent in A $\beta$  assembly through the detailed biostructural study of short A $\beta$  fragments. Many of these fragments undergo self-assembly and the resulting nanostructures remarkably resemble full length A $\beta$  (1-42). Previous collaboration between our group and the Lynn group focused on the 10-35 segment of the A $\beta$  (1-42) peptide. Molecular dynamics simulations on this beta sheet assembly showed that transient hydrogen bonds enforcing momentary planarity in alternate beta sheet regions of the peptide confer stability on the structures<sup>249</sup>.

In recent years the Lynn group has discovered that under specific conditions of solvent, ionic strength and pH, the 16-22 fragment (called A $\beta$  16-22 henceforth) of the A $\beta$  (1-42) amyloid peptide composed of the amino acids KLVFFAE forms very characteristic but distinctly different amyloid architectures<sup>250</sup>. This seven-residue is also of particular interest since four disease-causing mutations occur in this region<sup>251</sup>. In addition, unlike full-length amyloid A $\beta$  (16-22) is soluble and can be studied by spectroscopic techniques. In 40% acetonitrile-60% water solution at pH 6, A $\beta$  (16-22) forms helical fibers while at pH 2 it forms hollow nanotubes<sup>252</sup> (**Figure 6.3**). Measurements indicated the tubes to be double-walled with a wall thickness of 43 Å and a diameter of 520 Å. In case of fibers,

TEM reveals the presence of two 50 Å fibrils laterally associating to form twisted fibers with a maximum observed width of 100 Å. Characteristic powder x-ray reflections at 4.7 Å and 9.9 Å for both assemblies indicated the hydrogen-bonding and intersheet distance repeats respectively.



**Figure 6.3:** Structure of A $\beta$  (16-22) peptide fibers and nanotubes in 40% acetonitrile.

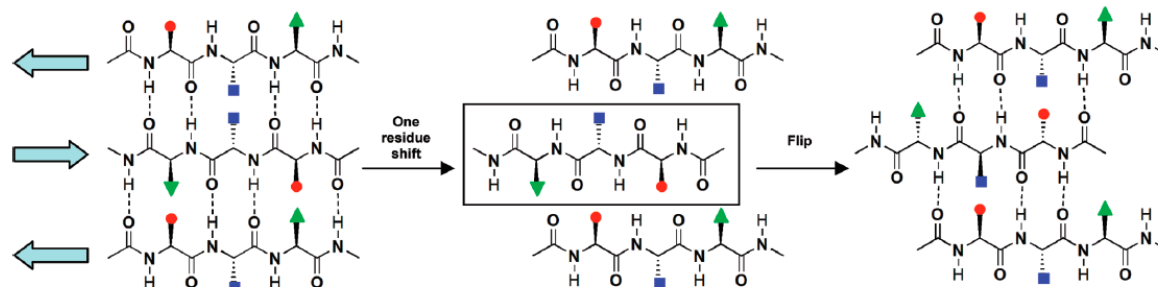
TEM micrograph of (a) fibers formed at pH 6 (inset shows twisted fiber dimer, scale bar= 50 nm) and (b) tubes at pH 2. Cryo-etch high-resolution SEM images of nanotubes with homogeneous diameters (scale bar= 250 nm) (c) and hollow cross-sections (d) (Figure and legend are from Lu et al.<sup>250</sup>)



Both assemblies show characteristic cross  $\beta$  structure and the corresponding experimental structural parameters. The only discernible difference between the monomers pertains to the ionization state of the terminal glutamates; negatively charged at pH 6 and unionized at pH 2. The drastic difference in structure corresponding to a rather simple difference in pH prompted us to explore the self-assembly of this peptide segment.

The Lynn group has used a battery of methods to characterize the A $\beta$  (16-22) assemblies. These include scanning electron and transmission electron microscopy for morphological characterization, small angle neutron scattering (SANS), small angle x-ray scattering (SAXS), solid-state NMR and isotope edited infrared spectroscopy (IR). Extensive studies using these techniques revealed fundamental differences between the  $\beta$ -sheet units comprising the tubes and fibers. The major features characterizing the two architectures lies in the orientation of the peptide strands with respect to each other, and the registry of the strands. Both assemblies posited the peptides as being anti-parallel, an arrangement deduced by isotope-edited IR measurements on  $^{13}\text{C}$  labeled peptides.

However, in case of the fibers, the peptides were all in-register while in case of the tubes, they were out of register by one residue<sup>252,253</sup>. The out-of-register nature of the peptides in the tubes necessitated the flipping of alternate strands to preserve hydrogen bonding, as shown in **Figure 6.4**. As demonstrated later, this particular geometric change has profound implications for the self-assembly of the two kinds of  $\beta$  sheets.



**Figure 6.4:** Flipping of central strand in a three-strand  $\beta$  sheet necessitated by a shift in register of the strand by one residue. (Figure from Liang et al.<sup>253</sup>)

To explore the factors affecting the variable self-assembly and resulting morphology under the two conditions, we decided to build computational models of both assemblies. At this point we did not think it pertinent to try to model the actual processes leading to self-assembly since the extended temporal characteristics of these processes located them beyond the scope and capabilities of current dynamical computational methods such as molecular dynamics that are still constrained to practically operate within time periods of several nanoseconds. In addition, the complex kinetics of self-assembly and the resulting lack of intermediate or nucleating structures limited an application of computational methods to modeling the process. Emphasis thus was put on the final structures themselves. Since the  $\beta$  sheet characteristics and the intrastrand and intersheet distances were already known, the goal instead was to produce a model that would capture these geometric particulars.

We decided to use molecular dynamics (MD) to model the structures. MD has often been used in the past to model amyloid self-assembly and has been generally validated as a

well-parameterized and benchmarked method for investigating proteins and peptides<sup>31</sup>. As stated above, our goal was not to use MD to model the dynamic self-assembly process but rather to construct models that agreed with the experimental constraints. For the MD simulations we decided to use the widely available program GROMACS (Groningen Machine for Chemical Simulations)<sup>254</sup>. GROMACS has been used extensively for simulation of biomolecules and benefits from the virtue of being faster than most other MD programs, especially on multiple processors. GROMACS uses the OPLS all-atom (AA)<sup>255</sup> and GROMOS force field for calculating interatomic interactions and uses the efficient Particle Mesh Ewald method for calculating long-range electrostatic interactions<sup>256,257</sup>. A unique algorithm to rapidly calculate non-bonded interactions that usually comprise the bulk of MD calculation time enables GROMACS to be, in the words of its creators, “Fast, Flexible and Free”<sup>254</sup>.

The calculations were initiated by Lynn lab postdoc Anil Mehta whose protocol was later modified and used. Simple system comprising  $\beta$  sheets were built in the Maestro interface (v9.1, Schrödinger). To explore the effect of size, symmetrical arrangements of strands and sheets, for instance 6 sheets with 6 strands each, were constructed. To avoid bias from edge effects, at least 5 strands in a sheet had to be maintained. Several arrangements including 5/5 and 8/8 were investigated. The results below are for the 5/5 assembly, although they did not differ in their essential details for other systems. A single  $\beta$  sheet was initially built in the desired in-register or one residue out-of-register orientation. As mentioned before, an out-of-register strand in a sheet needed to be flipped to achieve optimal hydrogen bonding. A single sheet thus constructed was first optimized using a

short force field minimization (OPLS 2005 or MMFFs) using the Macromodel<sup>122</sup> module in Maestro. The requisite number of such sheets was then stacked on top of each other at an approximate distance of 10 Å and further optimized in Macromodel. The system was thus set up for MD simulations.

MD simulations were carried out using GROMACS (v 3.2). The peptide assemblies were surrounded with a truncated octahedral box of about 25,000 SPC<sup>258</sup> water molecules. A truncated octahedron is essentially a cube with its corners truncated; a shape that eliminates extra volume of water molecules and this reduced computational time. For the peptides composing the fiber, the requisite number of positive sodium ions was added to compensate for the negatively charged glutamates and render the system electrically neutral. This system was then first subjected to a short force field minimization to eliminate bad contacts. After this a short 10 ps restrained MD simulation was run on it in which the peptide was restrained to let the water molecules equilibrate with the system.

The peptides were now ready for unrestrained MD simulations. These simulations were typically run for 1 ns. The time step employed was 2 fs, hydrogen bond lengths were constrained with the LINCS algorithm and the solute and solvent were coupled to temperature and pressure baths. After the simulation was over, RMSD plots were drawn for plotting the distances between stacked C $\alpha$ -C $\alpha$  atoms in two sequential beta sheets. As mentioned above, since the goal was to construct models agreeing with the experimental restraints, the part of the 1 ns trajectory corresponding to an average C $\alpha$ -C $\alpha$  distance of 9.9 Å was considered. Final models were generated by performing 50 ps steepest-descent

in vacuo minimizations over the last 800 ps of the trajectory, where the average intersheet distance was 9.9 Å.

The next step was to explore the factors responsible for the differential assembly of the peptides into fibers and tubes. As mentioned before, initial knowledge of the assemblies posits the peptides in the fibers anti-parallel in register. In this configuration at pH 6, the terminal ionized glutamate residue in one strand faces the terminal lysine residue in the adjacent strand, and one would expect a stabilizing salt bridge interaction between the two residues to manifest itself as a stabilizing factor. Because strands in the tube peptides are flipped and one residue out of register, they don't accrue the benefit of such an ionic interaction (**Figure 6.5**).



**Figure 6.5:** Schematic of two adjacent strands in a fiber assembly (left) and tube assembly (right). The proximity of the glutamate and lysine residues in the fiber is apparent (Figure from Mehta et al.<sup>252</sup>)

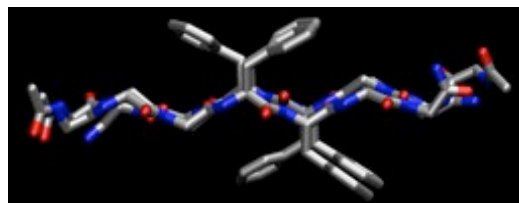
However, the role of salt bridges in stabilizing protein assemblies is not always obvious. Several studies in the past decade have shown that salt bridges can confer only marginal stability (1-2 kcal/mol) because of the penalty of desolvation, especially long-range desolvation, which their formation has to pay<sup>259</sup>. Nonetheless, protein charge ‘ladders’

and salt bridges have been postulated as elements of stability in both amyloid<sup>260</sup> and other protein structures<sup>261</sup> and small contributions from multiple such interactions can cumulatively add up to a substantial energetic gain. That the single salt bridge in each strand makes a discernible contribution to the fiber assembly was validated by studies with a A $\beta$  (16-22) mutant sequence KLVFFAL where the terminal glutamate was changed to a leucine. As hypothesized, this peptide did not self-assemble into fibers and formed tubes consisting of out-of-register  $\beta$  sheets at both pH 2 as well as pH 6, thus providing support for the contribution of the K-E salt bridge to fiber stability<sup>252</sup>.

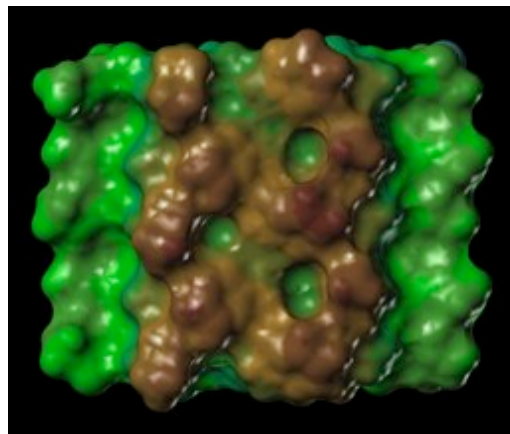
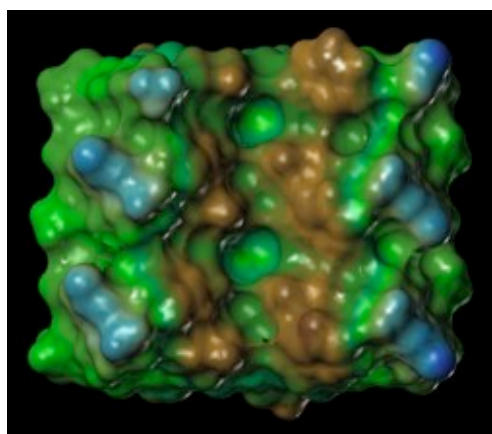
While the contribution of salt bridge interactions to fiber stability is readily contemplated, the factors responsible for tube assembly are not so apparent. Not only does the out-of-register nature of the tube peptide strands preclude the formation of a salt bridge as noted above, but the shift in registry also results in one fewer hydrogen bond compared to fiber sheets. Thus it would seem that beta sheet formation in the tube assembly would not be as facile as that in the fiber. However, observation of the surfaces formed by the tube and fiber peptide sheets provides strong hints of the differences between the two structures. As mentioned above, the principal difference between the tube and fiber peptide strands pertains to the flipping of adjacent strands in the tube that permit them to assemble one residue out-of-register. Further studies indicated that this geometric change leads to a profound difference in the lipophilic nature of the surfaces, the orientation in which the amino acids present themselves, and the resulting physical interactions between them.

For best realizing these differences, the central three-strand units in fiber and tube sheets were independently considered. Lipophilic potentials were mapped onto Connolly surfaces generated using a 1.4 Å probe in Sybyl (v 7.2). Illustrations of the top and bottom surfaces of tube and fiber sheets reveal the differences between them. It is also clear that these differences arise because of the dissimilar presentation of amino acids on the top and bottom surfaces of the two assemblies (**Figures 6.6 a, b and c**)

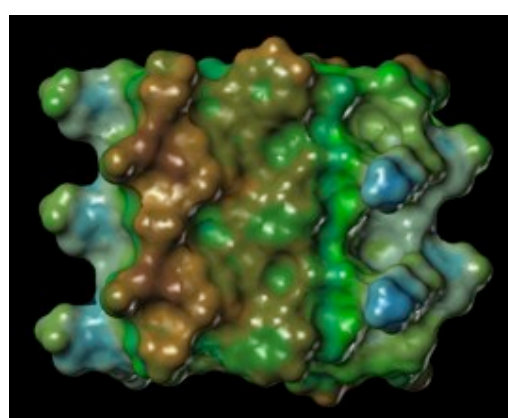
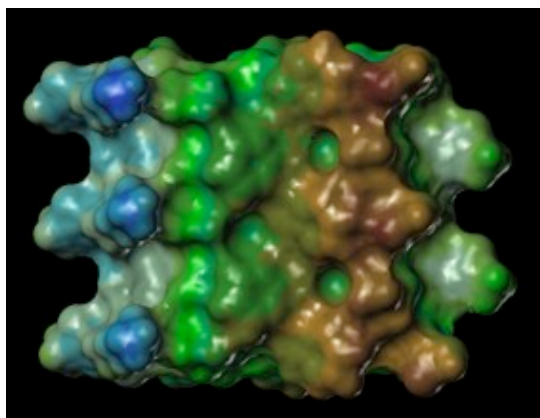
a.



b.



c.

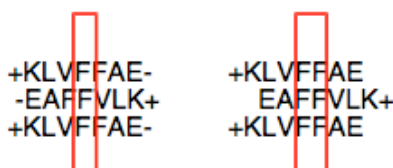




**Figure 6.6:** **a.** Edge view of three-strand beta sheet for fiber (left) and tube (right). **b.** Top (left) and bottom (right) views of lipophilic surfaces for three-strand fiber sheet. **c.** Top (left) and bottom (right) views of lipophilic surfaces for three-strand tube sheet. Blue denotes hydrophilic, brown denotes lipophilic and green denotes neutral. (Figure 6a. from Mehta et al.<sup>252</sup>)

**Figure 6.6a** illustrates that the top and bottom surfaces for fiber sheets are unsymmetrical compared to those for the tube sheets. These differences in presentation of the amino acid residues on the surface translate into the drastic difference in lipophilic surfaces for the two kinds of assemblies as shown in Figures 6b and 6c. The lipophilic surface for the tube sheet is similar and symmetrical compared to that for the fiber sheet. This allows the tube  $\beta$  sheets to stack in three potential arrangements; top-top, bottom-bottom, and top-bottom. On the other hand, it is energetically more favorable because of the complementarity of the top and bottom surfaces of the fiber sheet for them to preferentially stack as top-bottom units. Thus, once formed, lamination for tube sheets could be more rapid and more facile compared to fiber sheets, a fact reflected in the greater diameter of the tubes (520 Å) compared to the fibers (50 Å). Another factor that could aid in the favorable lamination relates to the stacking interactions of the central phenylalanines. The central phenylalanine in A $\beta$  (16-22) is a crucial residue that's key for self-assembly; sequences in which this residue is replaced by aliphatic or small lipophilic amino acids typically fail to display any organized structure formation. Stacking of Phe residues is also known to facilitate peptide-peptide interactions<sup>262,263</sup> and self-assembly in other systems and the energetic stabilization engendered by such stacking has now been studied both experimentally and by high-level quantum chemical methods<sup>264</sup>, thus

establishing the importance of aromatic stacking interactions in guiding self-assembly in both biology and materials science<sup>265</sup>. In case of the tube  $\beta$  sheets, an examination of the geometry clearly indicates that the one residue shifted registry reinforces potential stacking aromatic-aromatic interactions between two pairs of Phe residues per three stacked peptides, compared to one pair in the fiber beta sheet (**Figure 6.7**).



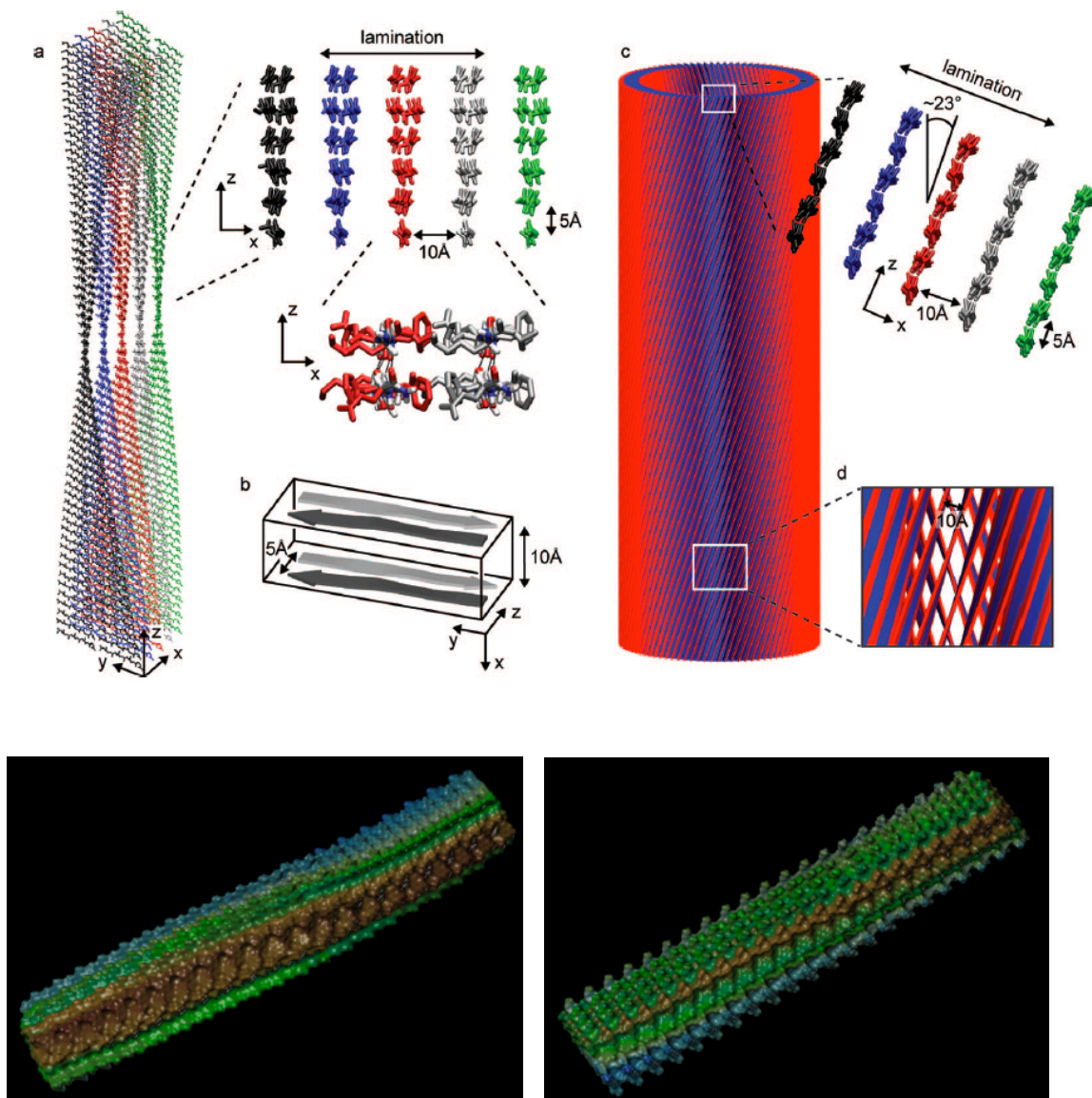
**Figure 6.7:** Differences in stacking between fiber peptides (left) and tube peptides (right)

The preferred orientation in these interacting pairs is seen to be the T-shaped or displaced T-shaped orientation. A battery of experimental and theoretical studies has located this configuration to be the most stable among three possible aromatic-aromatic configurations<sup>264</sup> (stacked, T-shaped, and parallel-displaced), although some studies have calculated it to be equienergetic to the parallel-displaced configuration<sup>264</sup>. The pronounced energetic stability of this configuration can be seen to contribute to the facility of the  $\beta$  sheet stacking, especially in the tube. Sequences in which Phe was replaced by Trp or Tyr did assemble, but only into fibers, indicated both the more problematic steric nature and lack of stacking proclivity in these residues<sup>252</sup>.

Other interactions contributing to sheet assembly include the displacement of water by the hydrophobic effect<sup>266</sup> as well as potential intersheet salt bridge interactions in the tube. The larger lipophilic surface area of the two surfaces in the tube  $\beta$  sheets is yet

another factor contributing to the lamination propensity of the sheets. The latter factor, namely intersheet K-E salt bridge interactions, have been shown to play a negligible role through MD simulations and through the observation (*vide supra*) that replacement of the terminal glutamate (E) in the sequence still led to tube formation.

Taken together, the geometric parameters from experimental data, MD simulations and an understanding of the factors inherent in  $\beta$  sheet self-assembly leads to models of the tube and fiber as shown in **Figure 6.8** below. The differential polarity of surfaces in the fiber sheets purportedly leads to twisting while the lipophilic complementarity and stacking interactions in the tube sheets results in an increased propensity of lamination.



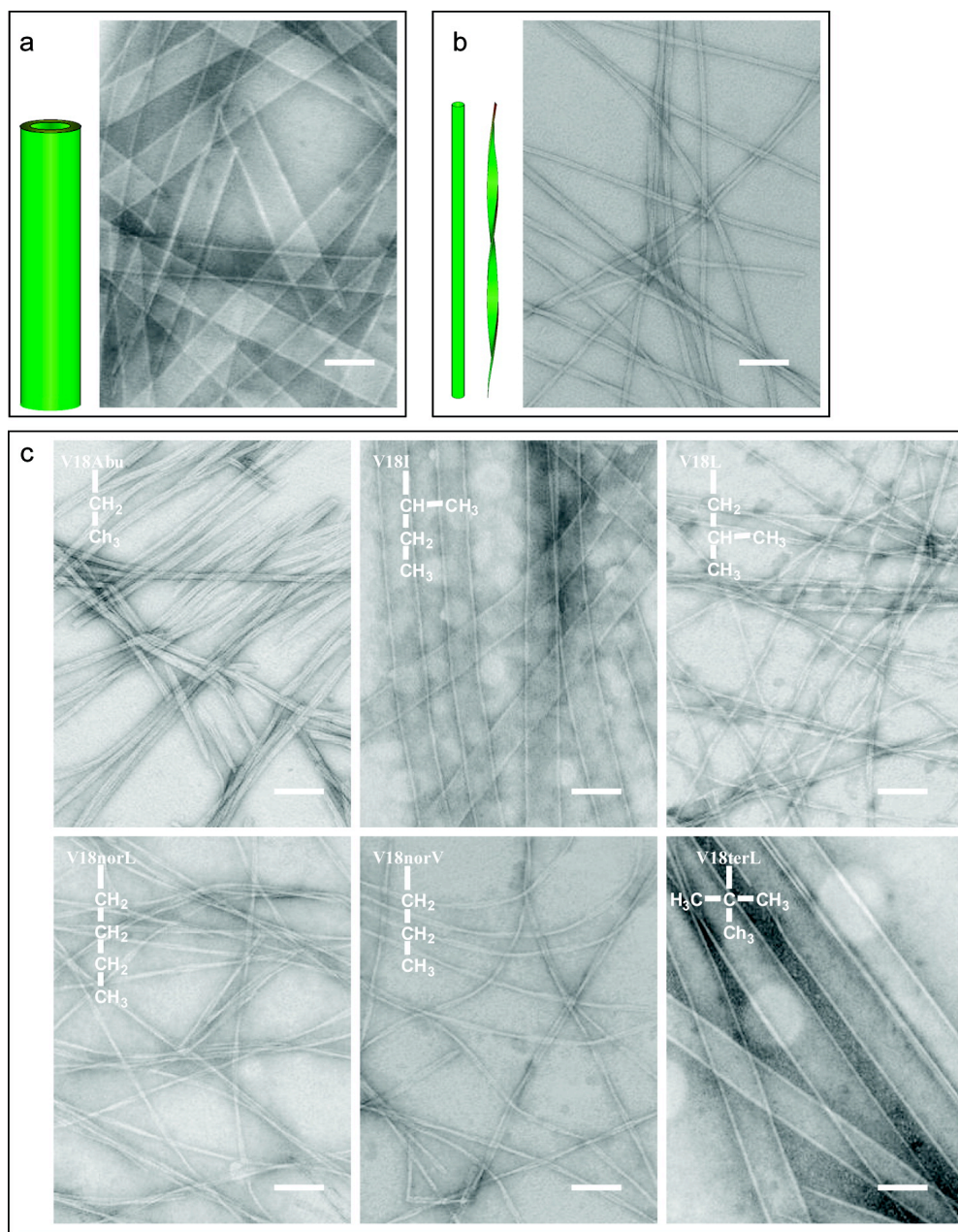
**Figure 6.8:** **Top a.** Twisted fiber assembly with hydrogen bonding between laminates running perpendicular to long axis. **B.** Unit cell, illustrating characteristic 5 Å hydrogen bonding and 10 Å intersheet distances. **c.** Tube assembly with hydrogen bonding parallel to long axis. **Bottom:** Differential polarity of two fiber surfaces illustrating lipophilic (brown), charged (blue) and neutral (green) surfaces. (Top figure from Mehta et al.<sup>252</sup>)

#### ***6.4: Cross-strand pairing and $\beta$ -sheet assembly***

Once we elucidated the factors responsible for lamination and supramolecular assembly of the  $\beta$  sheets, we wanted to investigate the interplay between such factors in governing structure formation in the  $\beta$  sheets themselves. Thus our goal was to try to elucidate the driving force that governs the coalescing of peptide strands to form a single  $\beta$  sheet in a particular configuration. Apart from hydrophobic surface area and salt bridge formation, another hitherto unexplored factor in formation of  $\beta$  sheets was the contribution of cross-strand pairing. Clearly the steric interactions between amino acid proximal to each other are important for modulating the resulting geometry. In this case we logically hypothesized that such size differences influencing cross-strand behavior would most significantly manifest themselves in peptide registry, a parameter which as mentioned above clearly leads to the different supramolecular assemblies of tubes and fibers. In the past, both experimental evidence<sup>267</sup> and bioinformatics analyses<sup>268,269</sup> have implicated cross-strand pairing as a contributor to  $\beta$  sheet assembly and stability in globular proteins. However, recent evidence suggests that formation of a single isolated  $\beta$  sheet might be regulated by hydrophobic surface burial events that dominate other  $\beta$  sheet propensity measures<sup>270</sup>. This difference suggests that the early steps in protein misfolding, events that have been debated for many years<sup>271,272</sup>, could be structurally probed in amyloid. We decided to perform both computational and experimental investigations to explore the effects of cross-strand pairing on peptide assembly<sup>253</sup>.

To test this hypothesis, Yan Liang from the Lynn group synthesized a series of point mutant sequences in which the V18 residue was changed to residues of variable steric characteristics and then performed extensive experiments on them to characterize their structure. In the antiparallel in-register arrangement, the V18 residue faces the F20 residue in the adjacent peptide. A series of single-point mutants constituting  $\beta$ -branched and other congeners changed the isopropyl side chain of the V18 residue to ethyl (Abu), *sec*-butyl (Ile), isobutyl (Leu), *n*-butyl (nor-Leu), *n*-propyl (nor-Val), and *t*-butyl (ter-Leu). The terL and Ile residues constitute  $\beta$ -branched amino acids while the others differ from Val by no more than a single  $-\text{CH}_2-$  group. The resulting assemblies at pH 2 and 6 were investigated with the set of experimental techniques cited above.

The experimental observations indicate the  $\beta$ -branched residues to have a distinct effect on peptide registry at the two pH values in 40% acetonitrile-60% water solutions. The registry seemed to be clearly influenced by the steric demand of the amino acid side chains. For instance, when the V18 position differed by no more than a single  $\text{CH}_2$  from valine, e.g., Abu, Leu, norL, and norV, only fibers formed under the acidic conditions (**Figure 6.9**). However, the  $\beta$ -branched amino acids Ile and terL directed nanotubes assembly, and the most sterically demanding terL congener directed nanotube assembly even under neutral conditions. X-ray measurements confirmed the characteristic amyloid pattern in these assemblies and isotope-edited IR measurements confirmed the peptide registry.



**Figure 6.9:** TEM of self-assembled structures: (a) A $\beta$  (16-22) nanotubes assembled under acidic conditions, (b) A $\beta$  (16-22) fibers assembled under neutral conditions, and (c) nanotubes or fibers formed by A $\beta$  (16-22) V18 congeners under acidic conditions; scale = 100 nm. Insets: The side chain structure of the residue substituted at the 18 position (Figure from Liang et al.<sup>253</sup>)

To shed light on these observations, molecular mechanics and molecular dynamics calculations were carried out on the sequences to evaluate the roles of steric and electrostatic interactions. MD simulations were carried out as before. Six copies of the seven-residue Ac-KLVFFAE-NH<sub>2</sub> peptide were combined twice graphically in Maestro v8.0 (Schrödinger) to produce two versions of the six-strand antiparallel  $\beta$  sheet with full hydrogen bonding as depicted in Scheme 1. The strands were organized either as in-register  $\beta$  sheets corresponding to fibers or as one-residue-shifted nanotubes. Five copies of the identical sheets were then stacked atop one another to provide five homogeneous laminates. To prepare for subsequent molecular dynamics simulations, the laminates were relieved of unfavorable torsions and steric contacts by energy minimization using the Truncated Newton Conjugate Gradient method and the GBSA/Water continuum solvation model. The relaxed peptide laminates were then enclosed in a truncated octahedral box and surrounded by 30,000-60,000 SPC water molecules in GROMACS v3.3. For peptides with positive charges, chloride ions were added to make the system neutral. The systems were prepared for MD simulations by performing initial energy minimizations on the aggregates for 10 ps using a steepest descent algorithm followed by solvent equilibration for 20 ps. Unrestrained MD was subsequently carried out for 2 ns at 300 K with a 2 fs time step using the OPLS 2005 force field. The resulting trajectories were viewed with VMD, and RMSD plots were generated using the xmgrace routine in GROMACS. Lipophilic potentials were mapped onto Connolly surfaces generated in the MOLCAD surface viewer with sphere radius of 1.4 Å in SYBYL 7.2. The MD results for the assemblies were similar as before with lipophilic surfaces mapped for the two registries. The MD simulations confirm that Lys-Glu intersheet salt bridges are abundant



in the in-register peptide sequences, but are not observed as intersheet stabilizing interactions in the one-residue out-of-register sheets. However, as demonstrated below, cross-strand pairing can sometimes be a more dominant factor in dictating the registry.

To further investigate the role that steric interactions play in the  $\beta$  sheet morphology, we used molecular mechanics to calculate the energies of the different in-register and one-residue-shifted assemblies. We chose molecular mechanics since force fields such as MMFFs and OPLS-2005 are well parameterized and primed to calculate sterics. On the other hand, force fields are also known to overemphasize electrostatic interactions primarily because of limitations in continuum solvation models. The importance of intrasheet salt bridges in the in-register  $\beta$  sheets has already been emphasized above. To eliminate effects arising from electrostatic salt bridge interactions in the present case and to try to tease out only the steric effects, we replaced the terminal lysines and glutamates with alanine. The resulting energy differences for the two different kinds of assemblies (in and out of register) were recorded.

For each peptide, the central three-strand  $\beta$  sheets were compared, either in-register or one-residue-shifted antiparallel registry, and the structures minimized with both OPLS 2005 and AMBER 94. When expressed as  $\Delta\Delta E$  ( $\Delta E_{\text{one-residue-shifted}} - \Delta E_{\text{in-register}}$ ), the in-register arrangement, which contains one additional H-bond, is lower in energy in each case. However, with  $\beta$  branched residues at 18, the energy difference is reduced significantly (**Table 6.1**), consistent with the  $\beta$ -branched residue at 18 preferring the one-residue-shifted registry. For example, the peptide with the non- $\beta$ -branched Leu at 18 has

the largest absolute  $\Delta\Delta E$ , about 4 kcal/mol. With the  $\beta$ -branched residue, Val or Ile, the  $\Delta\Delta E$ s are reduced to 2.5 kcal/mol, and with terL, the  $\Delta\Delta E$  is further reduced to 0.5 kcal/mol. These results are consistent with the experimental observation that peptides with  $\beta$ -branched residues at 18 significantly favor a one-residue-shifted registry and support side chain packing along the sheet face as a significant contributor to  $\beta$  sheet assembly and amyloid nucleation.

Peptides	OPLS 2005 (kcal/mol)	AMBER94 (kcal/mol)
Ac-ALVFFAA-NH <sub>2</sub>	-2.27	-2.43
Ac-ALIFFAA-NH <sub>2</sub>	-2.79	-3.24
Ac-AL- <b>terL</b> -FFAA-NH <sub>2</sub>	-0.61	-0.44
Ac-ALLFFAA-NH <sub>2</sub>	-3.93	-5.42

**Table 6.1:** Summary of the OPLS-05 and AMBER 94 energy difference between in-register and one-residue-shifted registries of A $\beta$ (16-22) and its V18 congeners. For each peptide, the simulation is done with 3-strand  $\beta$ -sheet in antiparallel in-register and antiparallel one-residue-shifted-register. The energy difference is calculated as  $\Delta\Delta E = \Delta E_{\text{in-register}} - \Delta E_{\text{one-residue-shifted}}$  for each peptide, in both force fields.

To determine the hydrophobic surface burial between non  $\beta$ -branched and  $\beta$ -branched amino acid side chains within the A $\beta$  (16-22) V18 congeners, their three-strand antiparallel in-register or one-residue-shifted  $\beta$  sheets were compared (**Table 6.2**). For

each peptide, the buried surface area was calculated by subtracting the solvent accessible surface area (SASA) value in the  $\beta$  sheet from its corresponding value in the random coil conformation<sup>270,273</sup>. The Val, Ile, or Leu congeners did not differ significantly in the mean fraction buried,  $f$ , which is an intrinsic measurement of the hydrophobicity<sup>273</sup>. More significantly, the peptides with Leu or Ile at 18 have the same number of atoms, and their difference in buried surface area upon forming the  $\beta$  sheet is less than 1% in both in-register and one-residue-shifted  $\beta$  sheets. Thus, unlike some previous studies, in this case buried surface area does not seem to be a strong indicator of the difference between the congener peptides that leads to tube or fiber assembly.

	Ac-ALVFFFAA-NH <sub>2</sub>	Ac-ALIFFFAA-NH <sub>2</sub>	Ac-ALLFFFAA-NH <sub>2</sub>
$A^0_T$	3990.3	4039.8	4076.1
$A^0_{B-B}$	1378.2	1362.6	1375.8
$A^0_{S-C}$	2612.1	2677.2	2700.3
Antiparallel One-residue-shifted $\beta$ -sheet			
$A_T$	1076.1	1107.6	1147
$A_{\text{phi}}$	583.4	583.6	593.3
$A_{\text{pho}}$	492.8	524	553.7
$\Delta A_T$	2914.2	2932.2	2929.1
$\Delta A_{\text{phi}}$	794.8	779.0	782.5
$\Delta A_{\text{pho}}$	2119.3	2153.2	2146.6
$f_T$	0.73	0.73	0.72
$f_{\text{phi}}$	0.58	0.57	0.57
$f_{\text{pho}}$	0.81	0.80	0.79
Antiparallel in-register $\beta$ -sheet			
$A_T$	1131.7	1100.6	1099.7
$A_{\text{phi}}$	527.9	545.5	539.4
$A_{\text{pho}}$	603.8	555.1	560.3
$\Delta A_T$	2858.6	2939.2	2976.4
$\Delta A_{\text{phi}}$	850.3	817.1	836.4
$\Delta A_{\text{pho}}$	2008.3	2122.1	2140.0
$f_T$	0.72	0.73	0.73
$f_{\text{phi}}$	0.62	0.60	0.61
$f_{\text{pho}}$	0.77	0.79	0.79

**Table 6.2:** Buried surface area of A $\beta$  (16-22) and its V18 congeners.

For each peptide, the simulation is done with a 3-strand  $\beta$  sheet. The unit is  $\text{\AA}^2$ .

$A^0_T$  is the total surface area obtained by adding the solvent accessible surface area

(SASA) of all seven residues in each peptide, and multiplied by 3 since this is a 3-strand  $\beta$  sheet. SASA value is from *Lesser & Rose, 1990*.

$A^0_{S-C}$  is the total SASA of the side chains. In our case, all the side chains are aliphatic or aromatic; therefore this value is defined as the hydrophobic surface area.

$A^0_{B-B}$  is the total SASA of the backbone, defined as hydrophilic area. Because our peptide

are capped at the both termini, the surface values of CH<sub>3</sub>, C=O are used as those in Ala, and the surface value of NH<sub>2</sub> is used as that in Lys.

The “ $A_{\text{phi}}$ ” and “ $A_{\text{pho}}$ ” are hydrophilic and hydrophobic surface area of each 3-strand  $\beta$  sheet calculated from *Maestro* (v 8.0). They are viewed as the hydrophilic or hydrophobic SASA in the folded state.

$$\Delta A_{\text{phi}} = A_{\text{B-B}}^0 - A_{\text{phi}}, \Delta A_{\text{pho}} = A_{\text{S-C}}^0 - A_{\text{pho}}$$

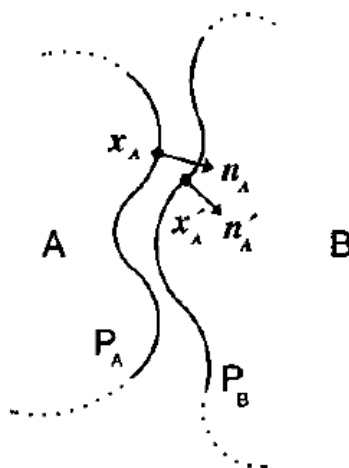
$f = \Delta A/A^0$ , which is defined as the mean fraction buried on the protein folding process (*Lesser & Rose, Protein: Structure, Function, and Genetics* 8, 6-13, 1990).

The above calculations and observations indicate that if at all buried hydrophobic surface area correlates to  $\beta$  sheet registry, it does so only marginally. In this particular case cross-strand pairing seems to be one of the primary factors dictating  $\beta$  sheet formation, even overriding salt bridge formation when bulky  $\beta$ -branched residues are present.

### ***6.5: Summary and future directions***

To summarize then, we tried to dissect factors operating in A $\beta$  (16-22) amyloid self-assembly into nanotubes and fibers under different physicochemical conditions. An attempt was made to individually investigate the contribution of steric, electrostatic and hydrophobic interactions in peptide  $\beta$  sheet self-assembly and lamination. In the future we would like to further explore the complementary ‘fit’ of the  $\beta$  beta sheets in the laminates. Quantitative or at least semi-quantitative parameters to differentiate between packing in fiber and tube  $\beta$  sheets would be valuable in knowing the extent of complementarity and to further shed light on the conditions prevalent in self-assembly.

To possibly do this, we have located a measure of surface complementarity used by other researchers that could accomplish this. This parameter called  $S_c$  was first suggested by Lawrence and Colman<sup>274</sup>. It measures the complementarity of two atomic Connolly surfaces generated by a probe by comparing the directions of unit vectors normal to the surfaces, originating from nearest points on the opposing surfaces (**Figure 6.10**).



**Figure 6.10:** Calculation of surface complementarity parameter  $S_c$  for two Connolly surfaces  $P_A$  and  $P_B$  involves the dot product of two unit vectors normal to the surfaces at points  $x_A$  and  $x_B$  ( $n_A \cdot n_B$ ) and the proximity of  $x_A$  and  $x_B$  ( $|x_A - x_B|$ ). The greater the dot product and the smaller the distance between the two points, the better the fit and complementarity between the two surfaces. (Figure adapted from Lawrence et al.)

$S_c$ , whose maximum value is 1.0 and minimum value is 0.0, depends in part on the proximity of the two points quantified by the absolute value of the distance between them, and the dot product of the two unit vectors, which approaches 1.0 in case of a perfect fit. In the past, this parameter has been used to quantify the unusually tight

packing of dry zipper-like interfaces in the Sup35 prion amyloid protein<sup>247</sup>. Typical values of  $S_c$  range from 0.76 for the rather tightly packed protease inhibitor proteins and their cognate proteases to the somewhat more ‘loosely’ packed antibody-inhibitor surfaces. At that point in time we did not have access to the CCP4 software that calculated this parameter, but it is hoped that in the future we will be able to accomplish this calculation for the different wild-type and mutant fiber and tube  $\beta$  sheets, thus providing a good measure of the packing propensities in the two kinds of assemblies that are influenced by various physicochemical factors.

The above investigations have helped to demonstrate the central biomolecular paradigm of dramatic differences in supramolecular structure resulting from small differences in molecular structure. This paradigm has had a long history in medicinal chemistry SAR where dramatic differences in biological activity can result from minor changes in chemical moieties<sup>13</sup>. It is striking that while medicinal chemists have a vast set of natural and synthetic chemical fragments to work from, nature has a more limited repertoire of twenty natural amino acids to fashion its structures. But as the astonishing diversity of protein structures and their varied functions demonstrate, nature has more than efficiently resolved the problem of generating complexity from simplicity by fine-tuning steric, electrostatic and other interactions. Much still remains to be known about amyloid structure, especially concerning the kinetics and the early nucleation events in self-assembly. However, the present work provides confidence that the self-assembly of naturally occurring amyloid in living systems could one day be comprehended at least partially through the understanding of subtle steric, electrostatic and lipophilic

modulation of the constituent units similar to that employed in the custom design of synthetic peptide fragments such as the ones above. There is no doubt that this understanding in turn will be useful in developing small molecules and other therapies that would exercise their action by interfering with such physicochemical interactions. And this understanding is critical if we want to gain insight into the treatment of diseases such as AD that are caused by protein misfolding and self-assembly.



# Chapter 7

## A Bird's-Eye View of Computer-Aided Drug Design

“A man would do nothing if he tried to do it so well  
that nobody would find fault with what he has done”

-John Henry Newman

“ Prediction is very difficult...especially about the future”

-Niels Bohr

### ***7.1: Drugs and rational drug discovery***

**N**atural substances have been used to treat mankind's diseases and ills since the dawn of humanity. The Middle Ages saw the use of exotic substances like sulfur and mercury to attempt to cure afflictions; most of these efforts resulted in detrimental side effects or death because of lack of knowledge of drug action. Quinine was isolated from the bark of the Cinchona tree and used for centuries to treat malaria. Salicylic acid was isolated from the Willow tree and was used for hundreds of years to treat fevers, knowledge that led to the discovery of Aspirin. The history of medicine has seen the use of substances ranging from arsenic to morphine, some of which are now known to be highly toxic or addictive.

The use of these substances reflected the state of medical knowledge of the times, when accidentally generated empirical data was the most valuable asset in the treatment of disease. Ancient physicians from Galen to Sushruta made major advances in our understanding of the human body and of medical therapies, but almost all of their knowledge was derived through patient and meticulously documented trial and error. A lack of knowledge of the scientific basis of disease meant that there were few systematic rational means of discovering new medicines, and serendipity and the traditional folk wisdom passed on through the centuries played the most important role in warding off disease.

This state of affairs continued till the 19<sup>th</sup> and 20<sup>th</sup> centuries when twin revolutions in biology and chemistry made it possible to discover drugs in a more logical manner.

Organic chemistry formally began in 1848 when Friedrich Wöhler found that he could synthesize urea from simple inorganic substances like ammonium cyanate, thus dispelling the belief that organic substances could only be synthesized by living organisms<sup>275</sup>. The further development of organic chemistry was orchestrated by the formulation of the structural theory in the late 19<sup>th</sup> century by Kekulé, Cooper, Kolbe, Perkin and others<sup>275</sup>. This framework made it possible to start to elucidate the precise arrangement of atoms in biologically active compounds. Knowledge of this arrangement in turn led to routes for synthesis of these molecules. These investigations also provided impetus to the synthesis of non-natural molecules of practical interest, sparking off the field of synthetic organic chemistry. However, while the power of synthetic organic chemistry later provided several novel drugs, the legacy of natural products is still prominent, and about half of the drugs currently on the market are either natural products or derived from natural products<sup>276</sup>.

Success in the application of chemistry to medicine was exemplified in the early 20<sup>th</sup> century by tentative investigations of what we currently call structure-activity relationships (SAR). Salvarsan, an arsenic compound used for treating syphilis, was perhaps the first example of a biologically active substance that had been improved by systematic investigation and modification. At the same time, chemists like Emil Fischer were instrumental in synthesizing further naturally occurring substances like carbohydrates and proteins, thus extending the scope of organic synthesis into biochemistry.

The revolution in structure determination initiated by physicists led to vastly improved synthesis and studies of bioactive substances. At this point, rational drug discovery began to take shape. Chemists working in tandem with biologists made hundreds of substances which were tested for their efficacy against various diseases. Knowledge from biological testing was in turn translated into modifications of the starting compounds. The first successful example of such rational efforts was the synthesis of sulfa drugs used to treat infections in the 1930s<sup>277</sup>. These compounds were the first effective antibiotics and were followed by the famous discovery, but this time serendipitous, of penicillin by Alexander Fleming in 1928<sup>278</sup>.

Rational drug discovery received a substantial impetus because of the post-World War 2 breakthroughs of structure determination by x-ray crystallography that revealed the structures of small molecules, proteins and DNA. The discovery of the structure of DNA in 1953 by Watson and Crick heralded the advent of molecular biology<sup>279</sup>. This landmark event led in succession to the elucidation of the genetic code and the transfer of genetic information from DNA to RNA that results in protein synthesis. The first structure determination of a protein -hemoglobin by Perutz<sup>280</sup> - was followed by the structure determination of several other proteins, some of which were pharmacologically important. Such advances and preceding ones by Pauling and others<sup>281</sup> led to the elucidation of common motifs in proteins such as alpha helices and beta sheets. The simultaneous growth of techniques in biological assaying and enzyme kinetics made it possible to monitor the binding of drugs to biomolecules. At the same time, better application of statistics and the standardization of double blind, controlled clinical trials

caused a fundamental change in the testing and approval of new medicines. A particularly noteworthy example of one of the first drugs discovered through rational investigations is cimetidine<sup>282</sup>, a drug for acid reflux that was for several years the best-selling drug in the world.

## ***7.2: Structure-based drug design and CADD***

As x-ray structures of protein-ligand complexes began to emerge in the 70s and 80s, rational drug discovery received enormous benefits. The development was also accompanied by High-Throughput Screening, an ability to screen thousands of ligands against a protein target to identify likely binders. These studies led to what today is known as “structure-based drug design” (SBDD)<sup>283</sup>. In SBDD, the structure of a protein bound to a ligand is used as a starting point for further modification and improvement of properties of the drug. While care has to be taken in order to fit the structure well to the electron density in the data<sup>284</sup>, well-resolved data can greatly help in identifying points of contact between the drug and the protein active site as well as the presence of special chemical moieties such as metals and cofactors. Water molecules identified in the active site can play crucial roles in bridging interactions between the protein and ligand<sup>285</sup>. Early examples of classes of drugs discovered using structure-based design include Captopril<sup>286</sup> (angiotensin-converting enzyme inhibitor- hypertension) and Trusopt<sup>287</sup> (carbonic anhydrase inhibitor- glaucoma) and recent examples include Aliskiren<sup>106</sup> (renin inhibitor- hypertension) and HIV protease inhibitors<sup>287</sup>. In this context, the discovery of BS-181 as a CDK7 inhibitor also constitutes an example of SBDD.

As SBDD progressed, another approach called ligand-based design (LBD) has also recently emerged. Obtaining x-ray structures of drugs bound to proteins is still a tricky endeavor, and one is often forced to proceed on the basis of the structure of an active compound alone. Techniques developed to tackle this problem involve QSAR (Quantitative Structure-Activity Relationships)<sup>288</sup> and pharmacophore construction in which the features essential for a particular ligand to bind to a certain protein are conjectured from affinity data for several similar and dissimilar molecules. Molecules based on the minimal set of interacting features are then synthesized and tested. However, since molecules can frequently adopt diverse conformations when binding to a protein, care has to be exercised in developing such hypotheses. In addition, it is relatively easy to be led astray by a high correlation between affinity data in the training set. It is paramount in such cases to remember the general discrepancy between correlation and causation, and overfitting of models can lead to both spurious correlations and absence of causation<sup>289</sup>. While LBD is more recent than SBDD, it has turned out to be valuable in certain cases. Noteworthy is a recent example where an inhibitor of NAADP was discovered by shape-based virtual screening<sup>290</sup> (vide infra)

As rational drug discovery progressed, software and hardware capacities of computers also grew exponentially, and CADD (Computer-Aided Drug Design) began to be increasingly applied to drug discovery. An effort was made to integrate CADD in the traditional chemistry and biology workflow and its principal development took place in the pharmaceutical industries, although academic groups were also instrumental in developing some capabilities<sup>291</sup>. The declining costs of memory and storage, increasing

processing power and facile computer graphics software put CADD within the grasp of relatively untrained computational chemists or experimental scientists. While the general verdict on the contribution of CADD to drug discovery is still forthcoming, many drugs currently on the market now include CADD as an important component of their discovery and development<sup>292</sup>. Many calculations that once were impractical because of constraints of time and computing power can now be routinely performed, some on a common desktop. Currently the use of CADD in drug design aims to address three principal problems, all of which are valuable to drug discovery.

### ***7.3: Virtual Screening:***

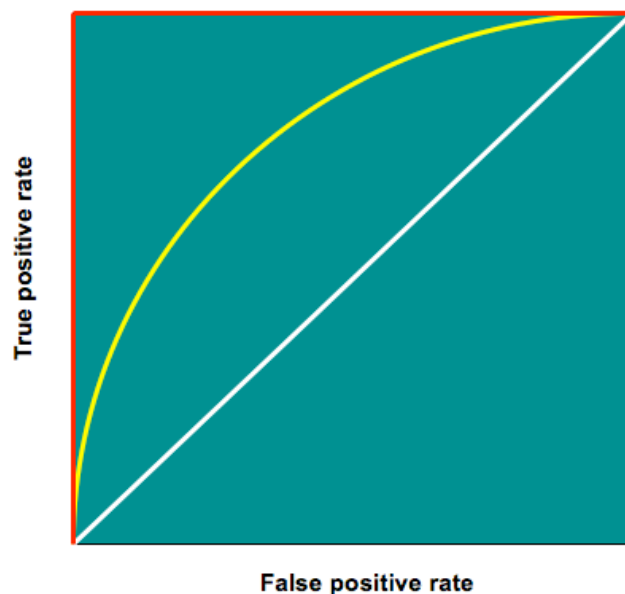
Virtual screening (VS) is defined by the ability to test thousands or millions of potential ligands against a protein, distinguish the actives from inactives and rank the 'true' binders in a certain top fraction. If validated, VS would serve as a valuable complement, if not substitute, for HTS and would save significant amounts of resources and time in HTS. Just like HTS, VS has to circumvent the problem of false positives and false negatives, the latter of which in some ways are more valuable since by definition they would not be identified. VS can be either structure-based or ligand-based. Both approaches have enjoyed partial success although recent studies have validated 3D ligand-based techniques in which ligand structures are compared to known active ligands by means of certain metrics as having a greater hit rate than structure-based techniques<sup>293</sup>. Virtual libraries of molecules such as DUD<sup>294</sup> (Directory of Useful Decoys) and ZINC<sup>295</sup> have been built to test the performance of several VS programs and compare them with each other. These libraries typically consist of a few actives and several thousand decoys, with

the goal being to rank the true actives above the true decoys using some metric.

Paramount in such retrospective assessment is an accurate method for evaluating the success and failure of these methods<sup>296,297</sup>. Until now ‘enrichment factors’ have mostly been used for this purpose<sup>297</sup>. The EF refers to the number of ‘true’ actives that rank in a certain top fraction (typically 1% or 10%) as a function of the screened database.

However the EF suffers from certain drawbacks, such as being dependent on the number of decoys in the dataset. To circumvent this problem, recent studies have suggested the use of the ROC (Receiver Operator Characteristic) curve, a graph that plots false positives vs. true positives<sup>297,298</sup> (**Figure 7.1**). The curve indicates what the false positive rate is for a given true positive rate and the measured variable is the Area Under the Curve (AUC). A completely random performance gives a straight line (AUC 0.5), while better performance results in a hyperbolic curve (AUC > 0.5).





**Figure 7.1:** ROC curve for three different VS scenarios. Completely random performance will give the straight white line (AUC 0.5), an ideal performance (no false positives and all true positives) will give the red line (AUC 1.0) and a good VS algorithm will produce the yellow curve ( $0.5 < \text{AUC} < 1.0$ )

Until now VS has provided limited evidence of success. Yet its capabilities are being improved and it has become a part of the computational chemist's standard repertoire. In some cases VS can provide more hits compared to HTS<sup>299</sup> and in others, VS at the very least provides a method to narrow down the number of compounds actually assayed<sup>300</sup>. As advances in general SBDD and LBD continue, the power of VS to identify true actives will undoubtedly increase.

#### **7.4: Pose-prediction:**

The second goal sought by computational chemists is to predict the binding orientation of a ligand in the binding pocket of a protein, a task that falls within the domain of SBDD.

This endeavor if successful will provide an enormous benefit in cases where crystal structures of protein-ligand complexes are not easily obtained. Since such cases are still very common, pose-prediction continues to be both a challenge as well as a valuable objective. There are two principal problems in pose prediction. The first one relates to the scoring of the poses obtained in order to identify the top-scoring pose as the ‘real’ pose; current docking programs are notorious for their scoring unreliability, certainly in an absolute sense and sometimes even in a relative sense. The problem of pose prediction ultimately is defined by the ability of an algorithm to find the global minimum orientation and conformation of a ligand on the potential energy surface (PES) generated by the protein active site<sup>301</sup>. As such it is susceptible to the common inadequacies inherent in comprehensively sampling a complex PES. Frequently however, as in the case of CDK7, past empirical data including knowledge of poses of known actives (roscovitine in this case) provides confidence about the pose of the unknown ligand.

Another serious problem in pose prediction is the inability of many current algorithms to adequately sample protein motion. X-ray structures provide only a static snapshot of ligand binding that may obscure considerable conformational changes in protein motifs. Molecular dynamics simulations followed by docking (‘ensemble docking’) have remedied this limitation to some extent<sup>302</sup>, induced-fit docking algorithms have now been included in programs such as GLIDE<sup>303</sup>, and complementary information from dynamical

NMR studies may help judicious selection between several protein poses. Yet simulating large-scale protein motions are still outside the domain of most MD simulations, although significant progress has been made in recent years<sup>304,305</sup>.

An example of how pose prediction can shed light on anomalous binding modes and possibly save the allocation of time and financial resources was experienced by the present author during his study of a paper detailing the development of inhibitors of the p38 MAP kinase<sup>306</sup>. In one instance the authors followed the SAR data in the absence of a crystal structure and observed contradictory changes in activity influenced by structural modifications. Crystallography on the protein ligand complex finally revealed an anomalous conformation of the ligand in which the oxygen of an amide at the 2 position of a thiophene was *cis* to the thiophene sulfur, when chemical intuition would have expected it to be *trans*. The crystal structure showed that an unfavorable interaction of a negatively charged glutamate with the sulfur in the more common *trans* conformation forced the sulfur to adopt the slightly unfavorable *cis* position with respect to the amide oxygen. Surprisingly this preference was seen in all top 5 GLIDE poses of the docked compound. This example indicates that at least in some cases pose prediction could serve as a valuable timesaving complement and possible alternative to crystallography.

### ***7.5: Binding affinity prediction***

The third goal is possibly the most challenging endeavor for computational chemistry. Rank-ordering ligands in terms of their binding affinity involves accurate scoring, which as noted above is a recalcitrant problem. The problem is a fundamental one since it really

involves calculating absolute free energies of protein ligand binding. The most accurate and sophisticated approaches for calculating these energies are the Free-Energy Perturbation (FEP)<sup>307</sup> or Thermodynamic Integration (TI) methods based on MD simulations and statistical thermodynamics. The methods involve ‘mutating’ one ligand to another in hundreds of thousands of infinitesimal steps and evaluating the binding enthalpy and entropy at every step. As of now, these techniques are some of the most computationally expensive techniques in the field. This problem typically limits their use only to evaluating free energy changes between ligand that differ little in structure. Therefore successful examples where they have found their greatest use involve cases where small substituents on aromatic rings are modified to evaluate changes in binding affinity<sup>308</sup>. However as computing power grows, these techniques will continue to find more applications in drug discovery.

Apart from these three goals, a major goal of computational science in drug discovery is to aid the later stages of drug development when pharmacokinetics (PK) and ADMET (Absorption Distribution Metabolism Excretion Toxicity) issues are key. Optimizing the binding affinity of a particular compound to a protein only results in an efficient ligand and not necessarily an efficient drug. Computational chemistry can make valuable contributions to these later developmental stages by trying to predict the relevant properties of ligands in the early stages, thus limiting the typically high attrition of drugs in the advanced phases. While much remains to be accomplished in this context, some progress has been made<sup>309</sup>. For example, the well-known Lipinski Rule of Five<sup>310</sup> provides a set of physicochemical properties necessary for drugs to have good

bioavailability and computational approaches are starting to help evaluate these properties during early stages. The QikProp program developed by Jorgensen et al. calculates properties like Caco-2 cell permeability, possible metabolites, % absorption in the GI tract and logP values<sup>311</sup>. Such programs are still largely empirical, depending on a large dataset of properties of known drugs for comparison and fitting.

### ***7.6: Models, computers and drug discovery***

In applying models to designing drugs and simulating their interactions with proteins, the most valuable lesson to remember is that these are models that are generated by computers. Models seldom mirror reality; in fact they often may succeed in spite of reality. Models are not usually designed to simulate reality but they are designed to produce results that agree with experiment. There are many approaches that produce such results. These approaches may not always encompass factors operating in real environments. In QSAR for instance, it has been shown that adding enough number of parameters to your model can lead to a good fit to the data with a high correlation coefficient. However the model may be overfitted; that is, it may seem to fit the known data very well but may fail to predict the unknown data, which is what it was designed to do<sup>289,312</sup>. In such cases, using more advanced statistical methods and using ‘bootstrapping’ (leaving out a part of the data and looking at the resulting fit to investigate whether that part of data is predicted) can lead to improvement in results<sup>312</sup>.

Models can also be used in spite of outliers. A high correlation coefficient of 0.85 that leads to acceptance of a model may nonetheless lead to one or two outliers. It then

becomes important to be aware of the physical anomaly which the outliers represent. The reason for this is clear. If the variable producing the outlier does not constitute a part of the model building, then applying the well-trained model to a system where that particular variable suddenly becomes dominant will result in a failure of the model. Such outliers, termed ‘black swans’, can prove extremely deleterious if their value is unusually high<sup>313</sup>. This phenomenon is known to operate in the field of financial engineering<sup>313</sup>. In modeling for instance, if the training set for a docking model consists of largely lipophilic protein active sites, then the model may fail to deliver cogent results if applied to a set of ligands binding to a protein that has an anomalously polar or charged active site. If the value of this protein is unusually high for a particular pharmaceutical project, an inability to predict its behavior under unforeseen circumstances may lead to valuable losses. Clearly in this case the physical variable, namely the polarity of the active site, was not taken into account in spite of the fact that the model delivered a high initial correlation merely because of the addition of a large number of parameters or descriptors, none of which was related in a significant way to the polarity of the binding pocket. The difference between correlation and causation is especially relevant in this respect. This hypothetical example illustrates one of the limitations of models iterated above; that they may not bear relationship to actual physical phenomena and may yet fit the data well enough because of various reasons to elicit confidence in their predictive ability.

In summary, models of the kind that are used in computational chemistry have to be carefully evaluated, especially in the context of practical applications like drug discovery where time and financial resources are valuable. Training the model on high-quality

datasets, reiterating the difference between correlation and causation and better application of statistics and bootstrapping can help to avert model failure.

In the end however, it is experiment that is of paramount importance for building the model. Inaccurate experimental data with uncertain error margins will undoubtedly hinder the success of every subsequent step in model building. To this end, generating, presenting and evaluating accurate experimental data is a responsibility that needs to be fulfilled by both computational chemists and experimentalists, and it is only a fruitful and synergistic alliance between the two groups that can help overcome the complex challenges in drug discovery.

*“In every investigation, in every extension of knowledge, we’re involved in action. And in every action we’re involved in choice. And in every choice we’re involved in a kind of loss, the loss of what we didn’t do. We find this in the simplest situations. . . . Meaning is always obtained at the cost of leaving things out. . . . In practical terms this means, of course, that our knowledge is always finite and never all encompassing. . . . This makes the world of ours an open world, a world without end. ”*      Robert Oppenheimer



## References

1. Feynman, R. P.; Leighton, R. B.; Sands, M. L. *Six easy pieces : essentials of physics, explained by its most brilliant teacher*; Helix Books: Reading, Mass., 1995.
2. Feynman, R. P.; Leighton, R. B.; Sands, M. L. *The Feynman lectures on physics*; Addison-Wesley Pub. Co.: Reading, Mass., 1963.
3. Nicolaou, K. C.; Snyder, S. A. *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 1012-44.
4. Laue, M. **1914**.
5. Bragg, W. L. **1922**.
6. Rhodes, G. *Crystallography made crystal clear : a guide for users of macromolecular models*; 2nd. ed.; Academic Press: San Diego, 2000.
7. Lambert, J. B.; Mazzola, E. P. *Nuclear Magnetic Resonance Spectroscopy: An Introduction to Principles, Applications, and Experimental Methods*; Prentice Hall, 2003.
8. Eliel, E. L.; Wilen, S. H. *Stereochemistry of Organic Compounds*; Wiley-Interscience, 1994.
9. Nicolaou, K. C.; Frederick, M. O. *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 5278-82.
10. Wuthrich, K. *Nat. Struct. Biol.* **2001**, *8*, 923-5.
11. Freeman, R. *Magnetic resonance in chemistry and medicine*; Oxford University Press: New York, 2003.
12. Patrick, G. L. *An Introduction to Medicinal Chemistry*; Oxford University Press, 2009.
13. Silverman, R. B. *The organic chemistry of drug design and drug action*; 2nd ed.; Elsevier Academic Press: Amsterdam ; Boston, 2004.
14. Perola, E.; Charifson, P. S. *J. Med. Chem.* **2004**, *47*, 2499-510.
15. Tirado-Rives, J.; Jorgensen, W. L. *J. Med. Chem.* **2006**, *49*, 5880-4.
16. Butler, K. T.; Luque, F. J.; Barril, X. *J. Comput. Chem.* **2009**, *30*, 601-610.
17. Snyder, J. P.; Nettles, J. H.; Cornett, B.; Downing, K. H.; Nogales, E. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 5312-6.
18. Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870-&.

19. Haasnoot, C. A. G.; Deleeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980**, *36*, 2783-2792.
20. Kurtkaya, S.; Barone, V.; Peralta, J. E.; Contreras, R. H.; Snyder, J. P. *J. Am. Chem. Soc.* **2002**, *124*, 9702-3.
21. Eliel, E. L.; Wilen, S. H.; Mander, L. N. *Stereochemistry of organic compounds*; Wiley & Sons: New York, 1994.
22. Kelso, M. J.; Hoang, H. N.; Appleton, T. G.; Fairlie, D. P. *J. Am. Chem. Soc.* **2000**, *122*, 10488-10489.
23. Smith, A. B.; LaMarche, M. J.; Falcone-Hindley, M. *Org. Lett.* **2001**, *3*, 695-698.
24. Burkert, U.; Allinger, N. L. *Molecular mechanics*; American Chemical Society: [Washington, D.C.], 1982.
25. Rappé, A. K.; Casewit, C. J. *Molecular mechanics across chemistry*; University Science Books: Sausalito, Calif., 1997.
26. Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379-4386.
27. Landis, C.; Allured, V. S. *J. Am. Chem. Soc.* **1991**, *113*, 9493-9499.
28. Mierke, D. F.; Kurz, M.; Kessler, H. *J. Am. Chem. Soc.* **1994**, *116*, 1042-1049.
29. Nikiforovich, G. V.; Kover, K. E.; Zhang, W. J.; Marshall, G. R. *J. Am. Chem. Soc.* **2000**, *122*, 3262-3273.
30. Cicero, D. O.; Barbato, G.; Bazzo, R. *J. Am. Chem. Soc.* **1995**, *117*, 1027-1033.
31. Adcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, *106*, 1589-1615.
32. Lakdawala, A. S.; Wang, M.; Nevins, N.; Liotta, D. C.; Rusinska-Roszak, D.; Lozynski, M.; Snyder, J. P. *BMC Biol.* **2001**, *1*, 2.
33. Nettles, J. H.; Li, H.; Cornett, B.; Krahn, J. M.; Snyder, J. P.; Downing, K. H. *Science* **2004**, *305*, 866-9.
34. Kingston, D. G. *J. Org. Chem.* **2008**, *73*, 3975-84.
35. Kelso, M. J.; Beyer, R. L.; Hoang, H. N.; Lakdawala, A. S.; Snyder, J. P.; Oliver, W. V.; Robertson, T. A.; Appleton, T. G.; Fairlie, D. P. *J. Am. Chem. Soc.* **2004**, *126*, 4828-4842.
36. Thepchatri, P.; Eliseo, T.; Cicero, D. O.; Myles, D.; Snyder, J. P. *J. Am. Chem. Soc.* **2007**, *129*, 3127-34.
37. Butler, K. T.; Luque, F. J.; Barril, X. *J. Comput. Chem.* **2009**, *30*, 601-10.

38. Barril, X.; Soliva, R. *Mol. Biosyst.* **2006**, *2*, 660-81.
39. Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. *J. Med. Chem.* **2006**, *49*, 5912-31.
40. Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. *J. Mol. Biol.* **2008**, *377*, 914-34.
41. Neuhaus, D.; Williamson, M. P. *The nuclear Overhauser effect in structural and conformational analysis*; 2nd ed.; Wiley: New York, 2000.
42. Snyder, J. A.; Mullins, J. M. *Cell. Biol. Int.* **1993**, *17*, 1075-84.
43. Jordan, M. A.; Wilson, L. *Nat. Rev. Cancer.* **2004**, *4*, 253-65.
44. Nicolaou, K. C.; Sasmal, P. K.; Rassias, G.; Reddy, M. V.; Altmann, K. H.; Wartmann, M.; O'Brate, A.; Giannakakou, P. *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 3515-20.
45. Smith, A. B.; Freeze, B. S. *Tetrahedron* **2008**, *64*, 261-298.
46. Florence, G. J.; Gardner, N. M.; Paterson, I. *Nat. Prod. Rep.* **2008**, *25*, 342-75.
47. Ravelli, R. B.; Gigant, B.; Curmi, P. A.; Jourdain, I.; Lachkar, S.; Sobel, A.; Knossow, M. *Nature* **2004**, *428*, 198-202.
48. Himes, R. H.; Kersey, R. N.; Heller-Bettinger, I.; Samson, F. E. *Cancer Res.* **1976**, *36*, 3798-802.
49. Denduluri, N.; Swain, S. M. *Expert. Opin. Investig. Drugs* **2008**, *17*, 423-35.
50. Buey, R. M.; Barasoain, I.; Jackson, E.; Meyer, A.; Giannakakou, P.; Paterson, I.; Mooberry, S.; Andreu, J. M.; Diaz, J. F. *Chem. Biol.* **2005**, *12*, 1269-79.
51. Thepchatrri, P.; Cicero, D. O.; Monteagudo, E.; Ghosh, A. K.; Cornett, B.; Weeks, E. R.; Snyder, J. P. *J. Am. Chem. Soc.* **2005**, *127*, 12838-46.
52. Gertsch, J.; Meier, S.; Muller, M.; Altmann, K. H. *ChemBioChem* **2009**, *10*, 166-75.
53. Ganesh, T.; Guza, R. C.; Bane, S.; Ravindra, R.; Shanker, N.; Lakdawala, A. S.; Snyder, J. P.; Kingston, D. G. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 10006-11.
54. Shanker, N.; Kingston, D. G.; Ganesh, T.; Yang, C.; Alcaraz, A. A.; Geballe, M. T.; Banerjee, A.; McGee, D.; Snyder, J. P.; Bane, S. *Biochemistry* **2007**, *46*, 11514-27.
55. Alcaraz, A. A.; Mehta, A. K.; Johnson, S. A.; Snyder, J. P. *J. Med. Chem.* **2006**, *49*, 2478-88.

56. Gunasekera, S. P.; Gunasekera, M.; Longley, R. E.; Schulte, G. K. *J. Org. Chem.* **1990**, *55*, 4912-4915.
57. Nerenberg, J. B.; Hung, D. T.; Somers, P. K.; Schreiber, S. L. *J. Am. Chem. Soc.* **1993**, *115*, 12621-12622.
58. Hung, D. T.; Chen, J.; Schreiber, S. L. *Chem. & Biol.* **1996**, *3*, 287-293.
59. Kowalski, R. J.; Giannakakou, P.; Gunasekera, S. P.; Longley, R. E.; Day, B. W.; Hamel, E. *Mol. Pharmacol.* **1997**, *52*, 613-622.
60. Klein, L. E.; Freeze, B. S.; Smith, A. B.; Horwitz, S. B. *Cell Cycle* **2005**, *4*, 501-507.
61. Xia, S.; Kenesky, C. S.; Rucker, P. V.; Smith, A. B., 3rd; Orr, G. A.; Horwitz, S. B. *Biochemistry* **2006**, *45*, 11762-75.
62. Madiraju, C.; Edler, M. C.; Hamel, E.; Raccor, B. S.; Balachandran, R.; Zhu, G. Y.; Giuliano, K. A.; Vogt, A.; Shin, Y. S.; Fournier, J. H.; Fukui, Y. H.; Bruckner, A. M.; Curran, D. P.; Day, B. W. *Biochemistry* **2005**, *44*, 15053-15063.
63. Raccor, B. S.; Vogt, A.; Sikorski, R. P.; Madiraju, C.; Balachandran, R.; Montgomery, K.; Shin, Y.; Fukui, Y.; Jung, W. H.; Curran, D. P.; Day, B. W. *Mol. Pharmacol.* **2008**, *73*, 718-726.
64. Shin, Y.; Fournier, J. H.; Fukui, Y.; Bruckner, A. M.; Curran, D. P. *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 4634-4637.
65. Paterson, I.; Britton, R.; Delgado, O.; Meyer, A.; Poullennec, K. G. *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 4629-4633.
66. Paterson, I.; Britton, R.; Delgado, O.; Wright, A. E. *Chem. Comm.* **2004**, 632-633.
67. Jung, W. H.; Harrison, C.; Shin, Y.; Fournier, J. H.; Balachandran, R.; Raccor, B. S.; Sikorski, R. P.; Vogt, A.; Curran, D. P.; Day, B. W. *J. Med. Chem.* **2007**, *50*, 2951-2966.
68. Paterson, I.; Naylor, G. J.; Wright, A. E. *Chem. Commun.* **2008**, 4628-4630.
69. Paterson, I.; Gardner, N. M.; Naylor, G. J. *Pure Appl. Chem.* **2009**, *81*, 169-180.
70. Jogalekar, A. S.; Krishnan, D.; Jung, W.-H.; Zhong, S.; Curran, D. P.; Snyder, J. P. **2009**.
71. Nevins, N.; Cicero, D.; Snyder, J. P. *J. Org. Chem.* **1999**, *64*, 3979-3986.
72. Paterson, I.; Britton, R.; Delgado, O.; Wright, A. E. *Chem. Comm.* **2004**, 632-3.
73. Canales, A.; Matesanz, R.; Gardner, N. M.; Andreu, J. M.; Paterson, I.; Diaz, J. F.; Jimenez-Barbero, J. *Chem. Eur. J.* **2008**, *14*, 7557-69.

74. Guimaraes, C. R. W.; Cardozo, M. J. *Chem. Inf. Model.* **2008**, *48*, 958-970.
75. Monteagudo, E.; Cicero, D. O.; Cornett, B.; Myles, D. C.; Snyder, J. P. *J. Am. Chem. Soc.* **2001**, *123*, 6929-30.
76. Sanchez-Pedregal, V. S.; Kubicek, K.; Meiler, J.; Lyothier, I.; Paterson, I.; Carlomagno, T. *Angew. Chem. Intl. Ed. Engl.* **2006**, *45*, 7388-7394.
77. Smith, A. B.; Beauchamp, T. J.; LaMarche, M. J.; Kaufman, M. D.; Qiu, Y.; Arimoto, H.; Jones, D. R.; Kobayashi, K. *J. Am. Chem. Soc.* **2000**, *122*, 8654-8664.
78. Jogalekar, A. S.; Kriel, F.; Burger, P.; Cicero, D.; Snyder, J. P. *J. Med. Chem.* **2009**.
79. Smith, A. B.; Freeze, B. S. *Tetrahedron* **2008**, *64*, 261-298.
80. Hoffmann, R. W.; Stahl, M.; Shopfer, U.; Frenking, G. *Chem. Eur. J.* **1998**, *4*, 559-566.
81. Hoffmann, R. W.; Stenkamp, D.; Trieselmann, T.; Gottlich, R. *Eur. J. Org. Chem.* **1999**, 2915-2917.
82. Kingston, D. G. I. 2008.
83. Hesse, M. *Alkaloids: nature's curse or blessing?*; Verlag Helvetica Chimica Acta ; Wiley-VCH: Zürich  
Weinheim ; New York, 2002.
84. Segal, M. S.; Goldstein, M. M.; Attinger, E. O. *Diseases of the Chest* **1957**, *32*, 305-309.
85. Anderson, J. T.; Ting, A. E.; Boozer, S.; Brunden, K. R.; Danzig, J.; Dent, T.; Harrington, J. J.; Murphy, S. M.; Perry, R.; Raber, A.; Rundlett, S. E.; Wang, J. M.; Wang, N.; Bennani, Y. L. *J. Med. Chem.* **2005**, *48*, 2756-2758.
86. Anderson, J. T.; Ting, A. E.; Boozer, S.; Brunden, K. R.; Crumrine, C.; Danzig, J.; Dent, T.; Faga, L.; Harrington, J. J.; Hodnick, W. F.; Murphy, S. M.; Pawlowski, G.; Perry, R.; Raber, A.; Rundlett, S. E.; Stricker-Krongrad, A.; Wang, J. M.; Bennani, Y. L. *J. Med. Chem.* **2005**, *48*, 7096-7098.
87. Aneja, R.; Dhiman, N.; Idnani, J.; Awasthi, A.; Arora, S. K.; Chandra, R.; Joshi, H. C. *Cancer Chemother. Pharmacol.* **2007**, *60*, 831-839.
88. Aneja, R.; Lopus, M.; Zhou, J.; Vangapandu, S. N.; Ghaleb, A.; Yao, J.; Nettles, J. H.; Zhou, B. F.; Gupta, M.; Panda, D.; Chandra, R.; Joshi, H. C. *Cancer Res.* **2006**, *66*, 3782-3791.
89. Biotechnology, C. 200.

90. Ye, K. Q.; Ke, Y.; Keshava, N.; Shanks, J.; Kapp, J. A.; Tekmal, R. R.; Petros, J.; Joshi, H. C. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1601-1606.
91. Snyder, J. P.; Nevins, N.; Cicero, D. O.; Jasen, J. *J. Am. Chem. Soc.* **2000**, *122*, 724-725.
92. Jogalekar, A. S.; Sun, A.; Min, J.; Thepchattri, P.; Lankin, D. C.; Snyder, J. P. **2009**.
93. Schrödinger: New York, New York, 2006.
94. Seetharaman, J.; Rajan, S. S. *Zeit. Fur Kristall.* **1995**, *210*, 111-113.
95. Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441-&.
96. Hehre, W. J. *A guidebook to molecular mechanics and quantum mechanics calculations*; Wavefunction Inc., 2003.
97. Dunitz, J. D. *Chem. Comm.* **2003**, 545-8.
98. Jakubke, H.-D.; Sewald, N. *Peptides from A to Z : a concise encyclopedia*; Wiley-VCH: Weinheim, 2008.
99. Wasco, W.; Tanzi, R. E. *Molecular mechanisms of dementia*; Humana Press: Totowa, N.J., 1997.
100. Jones, J.; 2nd ed.; Oxford University Press: Oxford ; New York, 2002.
101. Shafer, W. M.; Springer: Berlin ; New York, 2006.
102. Patrick, G. L. *An introduction to medicinal chemistry*; 3rd ed.; Oxford University Press: Oxford ; New York, 2005.
103. Otvos, L. *Peptide-based drug design*; Humana: Totowa, NJ, 2008.
104. Sadowsky, J. D.; Fairlie, W. D.; Hadley, E. B.; Lee, H. S.; Umezawa, N.; Nikolovska-Coleska, Z.; Wang, S.; Huang, D. C.; Tomita, Y.; Gellman, S. H. *J. Am. Chem. Soc.* **2007**, *129*, 139-54.
105. Cheng, R. P.; Gellman, S. H.; DeGrado, W. F. *Chem. Rev.* **2001**, *101*, 3219-32.
106. Wood, J. M.; Maibaum, J.; Rahuel, J.; Grutter, M. G.; Cohen, N. C.; Rasetti, V.; Ruger, H.; Goschke, R.; Stutz, S.; Fuhrer, W.; Schilling, W.; Rigollier, P.; Yamaguchi, Y.; Cumin, F.; Baum, H. P.; Schnell, C. R.; Herold, P.; Mah, R.; Jensen, C.; O'Brien, E.; Stanton, A.; Bedigian, M. P. *Biochem. Biophys. Res. Commun.* **2003**, *308*, 698-705.
107. Fowler, V. G., Jr.; Boucher, H. W.; Corey, G. R.; Abrutyn, E.; Karchmer, A. W.; Rupp, M. E.; Levine, D. P.; Chambers, H. F.; Tally, F. P.; Vigliani, G. A.; Cabell, C. H.; Link, A. S.; DeMeyer, I.; Filler, S. G.; Zervos, M.; Cook, P.; Parsonnet, J.; Bernstein, J.

- M.; Price, C. S.; Forrest, G. N.; Fatkenheuer, G.; Gareca, M.; Rehm, S. J.; Brodt, H. R.; Tice, A.; Cosgrove, S. E. *N. Engl. J. Med.* **2006**, *355*, 653-65.
108. Belshaw, P. J.; Schreiber, S. L. *J. Am. Chem. Soc.* **1997**, *119*, 1805-1806.
109. Werth, B. *The billion-dollar molecule : one company's quest for the perfect drug*; 1st Touchstone ed.; Simon & Schuster: New York, 1995.
110. Dentino, A. R.; Raj, P. A.; Bhandary, K. K.; Wilson, M. E.; Levine, M. J. *J. Biol. Chem.* **1991**, *266*, 18460-8.
111. Cavellierfrontin, F.; Pepe, G.; Verducci, J.; Siri, D.; Jacquier, R. *J. Am. Chem. Soc.* **1992**, *114*, 8885-8890.
112. Luy, B.; Frank, A.; Kessler, H. *Molecular drug properties : measurement and prediction*; Wiley-VCH: Weinheim, 2008.
113. Snyder, J. P.; Lakdawala, A. S.; Kelso, M. J. *J. Am. Chem. Soc.* **2003**, *125*, 632-3.
114. Zhang, X.; Nikiforovich, G. V.; Marshall, G. R. *J. Med. Chem.* **2007**, *50*, 2921-5.
115. Suarez, Y.; Gonzalez, L.; Cuadrado, A.; Berciano, M.; Lafarga, M.; Munoz, A. *Mol. Cancer Ther.* **2003**, *2*, 863-72.
116. Morino, T.; Masuda, A.; Yamada, M.; Nishimoto, M.; Nishikiori, T.; Saito, S.; Shimada, N. *J. Antibiot.* **1994**, *47*, 1341-3.
117. Hamaguchi, T.; Masuda, A.; Morino, T.; Osada, H. *Chem. Biol.* **1997**, *4*, 279-86.
118. Hamaguchi, T.; Sudo, T.; Osada, H. *FEBS Lett.* **1995**, *372*, 54-8.
119. Hamaguchi, T.; Takahashi, A.; Kagamizono, T.; Manaka, A.; Sato, M.; Osada, H. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2657-60.
120. Bisek, N.; Wetzel, S.; Arndt, H. D.; Waldmann, H. *Chem. Eur. J.* **2008**, *14*, 8847-60.
121. Jogalekar, A. S. *Biopolymers* **2009**.
122. Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440-467.
123. Kolossvary, I.; Guida, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 5011-5019.
124. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.
125. Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A.; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875-11882.

126. Glenn, M. P.; Kelso, M. J.; Tyndall, J. D. A.; Fairlie, D. P. *J. Am. Chem. Soc.* **2003**, *125*, 640-641.
127. Chatterjee, J.; Mierke, D. F.; Kessler, H. *Chem. Eur. J.* **2008**, *14*, 1508-1517.
128. UK, C. R. *Cancer Incidence Worldwide*, CRUK, 2008.
129. Society, A. C.; Society, A. C., Ed.; American Cancer Society: 2008.
130. Hanahan, D.; Weinberg, R. A. *Cell* **2000**, *100*, 57-70.
131. Futreal, P. A.; Coin, L.; Marshall, M.; Down, T.; Hubbard, T.; Wooster, R.; Rahman, N.; Stratton, M. R. *Nat. Rev. Drug. Disc.* **2004**, *4*, 177-183.
132. Hofseth, L. J.; Hussain, S. P.; Harris, C. C. *Trends. Pharmacol. Sci.* **2004**, *25*, 177-81.
133. Quintas-Cardama, A.; Kantarjian, H.; Cortes, J. *Nat. Rev. Drug. Disc.* **2007**, *6*, 834-48.
134. Bange, J.; Zwick, E.; Ullrich, A. *Nat. Med.* **2001**, *7*, 548-52.
135. Fischer, O. M.; Streit, S.; Hart, S.; Ullrich, A. *Curr. Opin. Chem. Biol.* **2003**, *7*, 490-5.
136. Chabner, B. A.; Roberts, T. G. *Nat. Rev. Canc.* **2005**, *5*, 65-72.
137. Remers, W. A. *Chemists at war : accounts of chemical research in the United States during World War II*; Clarice Publications: Tucson, Ariz., 2000.
138. Jamieson, E. R.; Lippard, S. J. *Chem. Rev.* **1999**, *99*, 2467-98.
139. Blackburn, G. M.; Gait, M. J. *Nucleic Acids in Chemistry and Biology*; Oxford University Press, 1996.
140. Araujo, R. P.; Liotta, L.; Petricoin, E. F. *Nat. Rev. Drug. Disc.* **2007**, *6*, 871-880.
141. Jang, S. H.; Wientjes, M. G.; Au, J. L. *J. Pharmacol. Exp. Ther.* **2001**, *298*, 1236-42.
142. Stryer, L.; Berg, J.; Tymoczko, J. *Biochemistry*; W. H. Freeman and Co., 2007.
143. Lehninger, A.; Cox, M. *Principles of Biochemistry*, 2004.
144. Moore, J. T.; Collins, J. L.; Pearce, K. H. *ChemMedChem* **2006**, *1*, 504-23.
145. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*; 5th ed., 2008.



146. Walsh, C. T. *Posttranslational Modification of Proteins*; Roberts and Co., 2005.
147. Westheimer, F. H. *Science* **1987**, *235*, 1173-8.
148. Hunter, T. *Cell* **1995**, *80*, 225-36.
149. Adams, J. A. *Chem. Rev.* **2001**, *101*, 2271-2290.
150. Tibes, R.; Trent, J.; Kurzrock, R. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 357-384.
151. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002**, *298*, 1912-1934.
152. Carpenter, G. *Bioessays* **2000**, *22*, 697-707.
153. Margutti, S.; Laufer, S. *ChemMedChem* **2007**, *1*, 1-26.
154. Marone, R.; Cmiljanovic, V.; Giese, B.; Wymann, M. P. *Biochim. Biophys. Acta* **2008**, *1784*, 159-185.
155. Sundstrom, T. J.; Anderson, A. C.; Wright, D. L. *Org. Biomol. Chem.* **2009**, *7*, 840-50.
156. Zhang, X.; Gureasko, J.; Shen, K.; Cole, P. A.; Kuriyan, J. *Cell* **2006**, *125*, 1137-1149.
157. Cohen, P. *Nat. Rev. Drug. Disc.* **2002**, *1*, 309-315.
158. Bridges, A. J. *Chem. Rev.* **2001**, *101*, 2541-2571.
159. Bikker, J. A.; Brooijmans, N.; Wissner, A.; Mansour, T. S. *J. Med. Chem.* **2009**, *52*, 1493-509.
160. Liao, J. J. *J. Med. Chem.* **2007**, *50*, 1-16.
161. Kasper, B.; Kallinowski, B.; Herrmann, T.; Lehnert, T.; Mechttersheimer, G.; Geer, T.; Ho, A. D.; Egerer, G. *Dig. Dis.* **2006**, *24*, 207-11.
162. Bikker, J. A.; Brooijmans, N.; Wissner, A.; Mansour, T. S. *J. Med. Chem.* **2009**, *6*, 1493-1509.
163. Apsel, B.; Blair, J. A.; Gonzalez, B.; Nazif, T. M.; Feldman, M. E.; Aizenstein, B.; Hoffman, R.; Williams, R. L.; Shokat, K. M.; Knight, Z. A. *Nat. Chem. Biol.* **2008**, *4*, 691-9.
164. Fabbro, D.; Ruetz, S.; Buchdunger, E.; Cowan-Jacob, S. W.; Fendrich, G.; Liebetanz, J.; Mestan, J.; O'Reilly, T.; Traxler, P.; Chaudhuri, B.; Fretz, H.;

- Zimmermann, J.; Meyer, T.; Caravatti, G.; Furet, P.; Manley, P. W. *Pharmacol. Ther.* **2002**, *93*, 79-98.
165. Prien, O. *ChemMedChem* **2006**, *1*, 1195-1196.
166. Shokat, K. *Chem. Biol.* **2005**, *12*, 621-637.
167. Toogood, P. *Med. Res. Rev.* **2001**, *21*, 487-498.
168. Foundation, N. 2001.
169. Morgan, D. O. *Annu. Rev. Cell Dev. Biol.* **1997**, *13*, 261-291.
170. Malumbres, M.; Barbacid, M. *Nat. Rev. Cancer* **2009**, *9*, 153-66.
171. Huse, M.; Kuriyan, J. *Cell* **2002**, *109*, 275-82.
172. Fisher, R. P. *J. Cell. Sci.* **2005**, *118*, 5171-80.
173. Harper, J. W.; Elledge, S. J. *Genes Dev.* **1998**, *12*, 285-9.
174. Lolli, G.; Johnson, L. N. *Cell Cycle* **2005**, *4*, 572-7.
175. Fu, M.; Wang, C.; Li, Z.; Sakamaki, T.; Pestell, R. G. *Endocrinology* **2004**, *145*, 5439-47.
176. Malumbres, M.; Barbacid, M. *Nat. Rev. Cancer.* **2001**, *1*, 222-31.
177. Knockaert, M.; Greengard, P.; Meijer, L. *Trends. Pharmacol. Sci.* **2002**, *23*, 417-25.
178. Senderowicz, A. M. *Oncogene* **2003**, *22*, 6609-20.
179. Benson, C.; White, J.; De Bono, J.; O'Donnell, A.; Raynaud, F.; Cruickshank, C.; McGrath, H.; Walton, M.; Workman, P.; Kaye, S.; Cassidy, J.; Gianella-Borradori, A.; Judson, I.; Twelves, C. *Br. J. Cancer* **2007**, *96*, 29-37.
180. Losiewicz, M. D.; Carlson, B. A.; Kaur, G.; Sausville, E. A.; Worland, P. J. *Biochem. Biophys. Res. Commun.* **1994**, *201*, 589-95.
181. McClue, S. J.; Blake, D.; Clarke, R.; Cowan, A.; Cummings, L.; Fischer, P. M.; MacKenzie, M.; Melville, J.; Stewart, K.; Wang, S.; Zhelev, N.; Zheleva, D.; Lane, D. P. *Int. J. Cancer.* **2002**, *102*, 463-8.
182. Raynaud, F. I.; Whittaker, S. R.; Fischer, P. M.; McClue, S.; Walton, M. I.; Barrie, S. E.; Garrett, M. D.; Rogers, P.; Clarke, S. J.; Kelland, L. R.; Valenti, M.; Brunton, L.; Eccles, S.; Lane, D. P.; Workman, P. *Clin. Cancer. Res.* **2005**, *11*, 4875-87.
183. Sherr, C. J.; Roberts, J. M. *Genes Dev.* **2004**, *18*, 2699-711.

184. Malumbres, M.; Barbacid, M. *Trends Biochem. Sci.* **2005**, *30*, 630-41.
185. Rossi, D. J.; Londesborough, A.; Korsisaari, N.; Pihlak, A.; Lehtonen, E.; Henkemeyer, M.; Makela, T. P. *Embo J.* **2001**, *20*, 2844-56.
186. Shapiro, G. I. *J. Clin. Oncol.* **2006**, *24*, 1770-83.
187. Fischer, P. M. *Cell Cycle* **2004**, *3*, 742-6.
188. Pallas, M.; Verdaguer, E.; Jorda, E. G.; Jimenez, A.; Canudas, A. M.; Camins, A. *Med. Hypotheses* **2005**, *64*, 120-3.
189. Roy, R.; Adamczewski, J. P.; Seroz, T.; Vermeulen, W.; Tassan, J. P.; Schaeffer, L.; Nigg, E. A.; Hoeijmakers, J. H.; Egly, J. M. *Cell* **1994**, *79*, 1093-101.
190. Serizawa, H.; Makela, T. P.; Conaway, J. W.; Conaway, R. C.; Weinberg, R. A.; Young, R. A. *Nature* **1995**, *374*, 280-2.
191. Rochette-Egly, C.; Adam, S.; Rossignol, M.; Egly, J. M.; Chambon, P. *Cell* **1997**, *90*, 97-107.
192. Chen, D.; Riedl, T.; Washbrook, E.; Pace, P. E.; Coombes, R. C.; Egly, J. M.; Ali, S. *Mol. Cell* **2000**, *6*, 127-137.
193. Larochelle, S.; Merrick, K. A.; Terret, M. E.; Wohlbold, L.; Barboza, N. M.; Zhang, C.; Shokat, K. M.; Jallepalli, P. V.; Fisher, R. P. *Mol. Cell* **2007**, *25*, 839-50.
194. Makela, T. P.; Parvin, J. D.; Kim, J.; Huber, L. J.; Sharp, P. A.; Weinberg, R. A. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 5174-8.
195. Korsisaari, N.; Rossi, D. J.; Paetau, A.; Charnay, P.; Henkemeyer, M.; Makela, T. P. *J. Cell Sci.* **2002**, *115*, 4275-84.
196. Ni, Z.; Schwartz, B. E.; Werner, J.; Suarez, J. R.; Lis, J. T. *Mol. Cell* **2004**, *13*, 55-65.
197. Chen, D.; Riedl, T.; Washbrook, E.; Pace, P. E.; Coombes, R. C.; Egly, J. M.; Ali, S. *Mol. Cell* **2000**, *6*, 127-37.
198. Jordan, V. C. *Br. J. Pharmacol.* **2006**, *147 Suppl 1*, S269-76.
199. Massarweh, S.; Osborne, C. K.; Creighton, C. J.; Qin, L.; Tsimelzon, A.; Huang, S.; Weiss, H.; Rimawi, M.; Schiff, R. *Cancer Res.* **2008**, *68*, 826-33.
200. Osborne, C. K.; Bardou, V.; Hopp, T. A.; Chamness, G. C.; Hilsenbeck, S. G.; Fuqua, S. A.; Wong, J.; Allred, D. C.; Clark, G. M.; Schiff, R. *J. Natl. Cancer Inst.* **2003**, *95*, 353-61.

201. MacCallum, D. E.; Melville, J.; Frame, S.; Watt, K.; Anderson, S.; Gianella-Borradori, A.; Lane, D. P.; Green, S. R. *Cancer Res.* **2005**, *65*, 5399-407.
202. Meijer, L.; Raymond, E. *Acc. Chem. Res.* **2003**, *36*, 417-25.
203. Heathcote, D.; Kroll, S.; Scheiper, B.; Jogalekar, A. S.; Brackow, J.; Siwicka, A.; Periyasamy, M.; Tolhurst, T. S.; Kanneganti, S. K.; Snyder, J. P.; Liotta, D.; Aboagye, E. O.; Barrett, A. G. M.; Ali, S.; Coombes, R. C. *Cancer Res.* **2009**.
204. G. D. Hawkins, D. J. G., G. C. Lynch, C. C. Chambers I. Rossi, J. W. Storer, J. Li, P. Winget, D. Rinaldi, D. A. Liotard, C. J. Cramer, and D. G. Truhlar; University of Minnesota: 1999.
205. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. *J. Med. Chem.* **2006**, *49*, 6177-96.
206. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739-49.
207. Lyne, P. D.; Lamb, M. L.; Saeh, J. C. *J. Med. Chem.* **2006**, *49*, 4805-8.
208. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J. Med. Chem.* **2004**, *47*, 1750-9.
209. Huang, N.; Shoichet, B. K. *J. Med. Chem.* **2008**, *51*, 4862-5.
210. Shan, Y.; Seeliger, M. A.; Eastwood, M. P.; Frank, F.; Xu, H.; Jensen, M. O.; Dror, R. O.; Kuriyan, J.; Shaw, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 139-44.
211. Zhang, C.; Shokat, K. M. *Tetrahedron* **2007**, *63*, 5832-5838.
212. Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 808-13.
213. Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2008**, *130*, 2817-31.
214. Dunitz, J. D. *Science* **1994**, *264*, 670-670.
215. Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Kolossvary, I.; Klepeis, J. L.; Layman, T.; Mcleavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y. B.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C. *Communications of the ACM.* **2008**, *51*, 91-97.
216. Kerns, E. H.; Di, L. *Drug-like properties : concepts, structure design and methods : from ADME to toxicity optimization*; Academic Press: Amsterdam ; Boston, 2008.

217. Tanzi, R. E.; Parson, A. B. *Decoding darkness : the search for the genetic causes of Alzheimer's disease*; Perseus Pub.: Cambridge, Mass., 2000.
218. Alzheimer's Disease Education & Referral Center (National Institute on Aging) In *NIH publication no. 03-3431*; Alzheimer's Disease Education & Referral Center: Silver Spring, MD, 2003.
219. Rhodes, R. *Deadly feasts : tracking the secrets of a terrifying new plague*; Simon & Schuster: [New York], 1997.
220. Aging, N. I. o.; (U.S.), N. I. o. H., Ed. 2008.
221. Thompson, C. A.; Spilsbury, K.; Hall, J.; Birks, Y.; Barnes, C.; Adamson, J. *BMC Geriatr.* **2007**, *7*, 18.
222. Gambassi, G.; Landi, F.; Lapane, K. L.; Sgadari, A.; Mor, V.; Bernabei, R. *J Neurol Neurosurg. Psychiatry* **1999**, *67*, 59-65.
223. Allegri, R. F.; Butman, J.; Arizaga, R. L.; Machnicki, G.; Serrano, C.; Taragano, F. E.; Sarasola, D.; Lon, L. *Int. Psychogeriatr* **2007**, *19*, 705-18.
224. Abbott, A. *Nature* **2008**, *456*, 161-4.
225. Brookmeyer, R.; Johnson, E.; Ziegler-Graham, K.; Arrighi, H. M. *Alzheimers & Dementia* **2007**, *3*, 186-191.
226. Lott, I. T.; Head, E. *Neurobiol. Aging* **2005**, *26*, 383-9.
227. Halliwell, B. *J. Neurochem.* **2006**, *97*, 1634-58.
228. Kumar-Singh, S.; De Jonghe, C.; Cruts, M.; Kleinert, R.; Wang, R.; Mercken, M.; De Strooper, B.; Vanderstichele, H.; Lofgren, A.; Vanderhoeven, I.; Backhovens, H.; Vanmechelen, E.; Kroisel, P. M.; Van Broeckhoven, C. *Human Mol. Gen.* **2000**, *9*, 2589-2598.
229. Hardy, J.; Allsop, D. *Trends. Pharmacol. Sci.* **1991**, *12*, 383-8.
230. Mudher, A.; Lovestone, S. *Trends Neurosci.* **2002**, *25*, 22-6.
231. Wenk, G. L. *J. Clin. Psychiatry* **2003**, *64 Suppl 9*, 7-10.
232. Wolfe, M. S. *Nat. Rev. Drug. Disc.* **2002**, *1*, 859-66.
233. Shankar, G. M.; Li, S.; Mehta, T. H.; Garcia-Munoz, A.; Shepardson, N. E.; Smith, I.; Brett, F. M.; Farrell, M. A.; Rowan, M. J.; Lemere, C. A.; Regan, C. M.; Walsh, D. M.; Sabatini, B. L.; Selkoe, D. J. *Nat. Med.* **2008**, *14*, 837-42.

234. Malaplate-Armand, C.; Florent-Bechard, S.; Youssef, I.; Koziel, V.; Sponne, I.; Kriem, B.; Leininger-Muller, B.; Olivier, J. L.; Oster, T.; Pillot, T. *Neurobiol. Dis.* **2006**, *23*, 178-89.
235. Koffie, R. M.; Meyer-Luehmann, M.; Hashimoto, T.; Adams, K. W.; Mielke, M. L.; Garcia-Alloza, M.; Micheva, K. D.; Smith, S. J.; Kim, M. L.; Lee, V. M.; Hyman, B. T.; Spires-Jones, T. L. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 4012-7.
236. Mazanetz, M. P.; Fischer, P. M. *Nat. Rev. Drug. Disc.* **2007**, *6*, 464-79.
237. Kukar, T. L.; Ladd, T. B.; Bann, M. A.; Fraering, P. C.; Narlawar, R.; Maharvi, G. M.; Healy, B.; Chapman, R.; Welzel, A. T.; Price, R. W.; Moore, B.; Rangachari, V.; Cusack, B.; Eriksen, J.; Jansen-West, K.; Verbeeck, C.; Yager, D.; Eckman, C.; Ye, W.; Sagi, S.; Cottrell, B. A.; Torpey, J.; Rosenberry, T. L.; Fauq, A.; Wolfe, M. S.; Schmidt, B.; Walsh, D. M.; Koo, E. H.; Golde, T. E. *Nature* **2008**, *453*, 925-9.
238. Melnikova, I. *Nat. Rev. Drug. Disc.* **2007**, *6*, 341-2.
239. Frydman-Marom, A.; Rechter, M.; Shefler, I.; Bram, Y.; Shalev, D. E.; Gazit, E. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 1981-6.
240. Stains, C. I.; Mondal, K.; Ghosh, I. *ChemMedChem* **2007**, *2*, 1674-92.
241. Eriksen, J. L.; Sagi, S. A.; Smith, T. E.; Weggen, S.; Das, P.; McLendon, D. C.; Ozols, V. V.; Jessing, K. W.; Zavitz, K. H.; Koo, E. H.; Golde, T. E. *J. Clin. Invest.* **2003**, *112*, 440-9.
242. Petkova, A. T.; Yau, W. M.; Tycko, R. *Biochemistry* **2006**, *45*, 498-512.
243. Eanes, E. D.; Glenner, G. G. *J. Histochem. Cytochem.* **1968**, *16*, 673-7.
244. Perutz, M. F.; Finch, J. T.; Berriman, J.; Lesk, A. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5591-5.
245. Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333-66.
246. Chiti, F.; Dobson, C. M. *Nat. Chem. Biol.* **2009**, *5*, 15-22.
247. Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riek, C.; Grothe, R.; Eisenberg, D. *Nature* **2005**, *435*, 773-8.
248. Vestergaard, B.; Groenning, M.; Roessle, M.; Kastrup, J. S.; van de Weert, M.; Flink, J. M.; Frokjaer, S.; Gajhede, M.; Svergun, D. I. *PLoS Biol.* **2007**, *5*, e134.
249. Lakdawala, A. S.; Morgan, D. M.; Liotta, D. C.; Lynn, D. G.; Snyder, J. P. *J. Am. Chem. Soc.* **2002**, *124*, 15150-1.
250. Lu, K.; Jacob, J.; Thiyagarajan, P.; Conticello, V. P.; Lynn, D. G. *J. Am. Chem. Soc.* **2003**, *125*, 6391-3.

251. Santini, S.; Wei, G.; Mousseau, N.; Derreumaux, P. *Structure* **2004**, *12*, 1245-55.
252. Mehta, A. K.; Lu, K.; Childers, W. S.; Liang, Y.; Dublin, S. N.; Dong, J.; Snyder, J. P.; Pingali, S. V.; Thiyagarajan, P.; Lynn, D. G. *J. Am. Chem. Soc.* **2008**, *130*, 9829-35.
253. Liang, Y.; Pingali, S. V.; Jogalekar, A. S.; Snyder, J. P.; Thiyagarajan, P.; Lynn, D. G. *Biochemistry* **2008**, *47*, 10018-10026.
254. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701-18.
255. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474-6487.
256. Leach, A. R. *Molecular modelling : principles and applications*; Longman: Harlow, England, 2001.
257. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
258. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermolecular forces : proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry held in Jerusalem, Israel, April 13-16, 1981*; D. Reidel ; Sold and distributed in the U.S.A. and Canada by Kluwer Boston: Dordrecht, Holland ; Boston, U.S.A. Hingham, MA, 1981.
259. Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1994**, *3*, 211-226.
260. Tarus, B.; Straub, J. E.; Thirumalai, D. *J. Am. Chem. Soc.* **2006**, *128*, 16159-16168.
261. Carbeck, J. D.; Colton, I. J.; Gao, J. M.; Whitesides, G. M. *Acc. Chem. Res.* **1998**, *31*, 343-350.
262. Mahalakshmi, R.; Raghothama, S.; Balaram, P. *J. Am. Chem. Soc.* **2006**, *128*, 1125-1138.
263. Aravinda, S.; Shamala, N.; Das, C.; Sriranjini, A.; Karle, I. L.; Balaram, P. *J Am Chem Soc* **2003**, *125*, 5308-15.
264. Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656-68.
265. Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. *J. Chem. Soc. Perk. Trans. 2* **2001**, 651-669.
266. Dill, K. A.; Truskett, T. M.; Vlachy, V.; Hribar-Lee, B. *Annu. Rev. Biophys. Biomol. Struct* **2005**, *34*, 173-99.
267. Smith, C. K.; Regan, L. *Science* **1995**, *270*, 980-2.

268. Wouters, M. A.; Curmi, P. M. *Proteins* **1995**, *22*, 119-31.
269. Hutchinson, E. G.; Sessions, R. B.; Thornton, J. M.; Woolfson, D. N. *Protein Sci.* **1998**, *7*, 2287-300.
270. Yan, S.; Gawlak, G.; Makabe, K.; Tereshko, V.; Koide, A.; Koide, S. *J. Mol. Biol.* **2007**, *368*, 230-43.
271. Wetzel, R. *Acc. Chem. Res.* **2006**, *39*, 671-9.
272. Westermark, P. *FEBS J.* **2005**, *272*, 5942-9.
273. Lesser, G. J.; Rose, G. D. *Proteins* **1990**, *8*, 6-13.
274. Lawrence, M. C.; Colman, P. M. *J. Mol. Biol.* **1993**, *234*, 946-50.
275. Berson, J. A. *Chemical creativity : ideas from the work of Woodward, Hückel, Meerwein and others*; 1st ed.; Wiley-VCH: Weinheim ; Chichester, 1999.
276. Paterson, I.; Anderson, E. A. *Science* **2005**, *310*, 451-3.
277. Hager, T. *The demon under the microscope : from battlefield hospitals to Nazi labs, one doctor's heroic search for the world's first miracle drug*; 1st ed.; Harmony Books: New York, 2006.
278. Macfarlane, G. *Alexander Fleming, the man and the myth*; Oxford University Press: Oxford [Oxfordshire] ; New York, 1985.
279. Judson, H. F. *The eighth day of creation : makers of the revolution in biology*; Expanded ed.; CSHL Press: Plainview, N.Y., 1996.
280. Ferry, G. *Max Perutz and the secret of life*; Cold Spring Harbor Laboratory Press: New York, 2008.
281. Hager, T. *Linus Pauling and the chemistry of life*; Oxford University Press: New York, 1998.
282. Black, J. *Annu. Rev. Pharmacol. Toxicol.* **1996**, *36*, 1-33.
283. Jhoti, H.; Leach, A. R. *Structure-based drug discovery*; Springer: Dordrecht, 2007.
284. Davis, A. M.; Teague, S. J.; Kleywegt, G. J. *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 2718-36.
285. Ball, P. *Chem. Rev.* **2008**, *108*, 74-108.
286. Smith, C. G.; Vane, J. R. *FASEB J.* **2003**, *17*, 788-9.
287. Kubinyi, H. *J. Recept. Signal Transduct. Res.* **1999**, *19*, 15-39.



288. Hansch, C.; Leo, A.; Hoekman, D. H. *Exploring QSAR*; American Chemical Society: Washington, DC, 1995.
289. Doweyko, A. M. *J. Comput. Aided Mol. Des.* **2008**, *22*, 81-9.
290. Naylor, E.; Arredouani, A.; Vasudevan, S. R.; Lewis, A. M.; Parkesh, R.; Mizote, A.; Rosen, D.; Thomas, J. M.; Izumi, M.; Ganesan, A.; Galione, A.; Churchill, G. C. *Nat. Chem. Biol.* **2009**, *5*, 220-6.
291. Snyder, J. P. *Med. Res. Rev.* **1991**, *11*, 641-62.
292. Jorgensen, W. L. *Science* **2004**, *303*, 1813-8.
293. McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J. F.; Cornell, W. D. *J. Chem. Inf. Model.* **2007**, *47*, 1504-19.
294. Huang, N.; Shoichet, B. K.; Irwin, J. J. *J. Med. Chem.* **2006**, *49*, 6789-801.
295. Irwin, J. J.; Shoichet, B. K. *J. Chem. Inf. Model.* **2005**, *45*, 177-82.
296. Jain, A. N.; Nicholls, A. *J. Comput. Aided Mol. Des.* **2008**, *22*, 133-9.
297. Hawkins, P. C.; Warren, G. L.; Skillman, A. G.; Nicholls, A. *J. Comput. Aided Mol. Des.* **2008**, *22*, 179-90.
298. Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. *J. Med. Chem.* **2005**, *48*, 2534-47.
299. Babaoglu, K.; Simeonov, A.; Irwin, J. J.; Nelson, M. E.; Feng, B.; Thomas, C. J.; Cancian, L.; Costi, M. P.; Maltby, D. A.; Jadhav, A.; Inglese, J.; Austin, C. P.; Shoichet, B. K. *J. Med. Chem.* **2008**, *51*, 2502-11.
300. Peach, M. L.; Tan, N.; Choyke, S. J.; Giubellino, A.; Athauda, G.; Burke, T. R.; Nicklaus, M. C.; Bottaro, D. P. *J. Med. Chem.* **2009**.
301. Jain, A. N. *J. Comput. Aided Mol. Des.* **2008**, *22*, 201-12.
302. Rao, S.; Sanschagrín, P. C.; Greenwood, J. R.; Repasky, M. P.; Sherman, W.; Farid, R. *J. Comput. Aided Mol. Des.* **2008**, *22*, 621-7.
303. Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. *J. Med. Chem.* **2006**, *49*, 534-53.
304. Shan, Y.; Seeliger, M. A.; Eastwood, M. P.; Frank, F.; Xu, H.; Jensen, M. O.; Dror, R. O.; Kuriyan, J.; Shaw, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 139-44.
305. Jensen, M. O.; Dror, R. O.; Xu, H.; Borhani, D. W.; Arkin, I. T.; Eastwood, M. P.; Shaw, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14430-5.

306. Goldberg, D. R.; Hao, M. H.; Qian, K. C.; Swinamer, A. D.; Gao, D. A.; Xiong, Z.; Sarko, C.; Berry, A.; Lord, J.; Magolda, R. L.; Fadra, T.; Kroe, R. R.; Kukulka, A.; Madwed, J. B.; Martin, L.; Pargellis, C.; Skow, D.; Song, J. J.; Tan, Z.; Torcellini, C. A.; Zimmitti, C. S.; Yee, N. K.; Moss, N. *J. Med. Chem.* **2007**, *50*, 4016-26.
307. Jorgensen, W. L.; Thomas, L. L. *J. Chem. Theor. Comp.* **2008**, *4*, 869-876.
308. Zeevaart, J. G.; Wang, L. G.; Thakur, V. V.; Leung, C. S.; Tirado-Rives, J.; Bailey, C. M.; Domaoal, R. A.; Anderson, K. S.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2008**, *130*, 9492-9499.
309. Martin, Y. C. *J. Med. Chem.* **2005**, *48*, 3164-70.
310. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug. Del. Rev.* **1997**, *23*, 3-25.
311. Ioakimidis, L.; Thoukydidis, L.; Mirza, A.; Naeem, S.; Reynisson, J. *Qsar & Comb. Sci.* **2008**, *27*, 445-456.
312. Hawkins, D. M. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1-12.
313. Taleb, N. *The black swan : the impact of the highly improbable*; 1st ed.; Random House: New York, 2007.
-