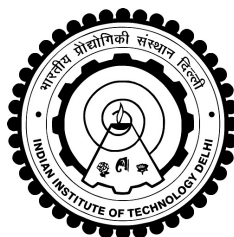


**THERMODYNAMICS, KINETICS, AND SOLUTION  
STRUCTURE ENSEMBLE OF THERAPEUTIC  
PROTEINS USING METADYNAMICS**

**RICHA SINGH**



**DEPARTMENT OF CHEMICAL ENGINEERING  
INDIAN INSTITUTE OF TECHNOLOGY DELHI**

**OCTOBER 2019**



**THERMODYNAMICS, KINETICS, AND SOLUTION  
STRUCTURE ENSEMBLE OF THERAPEUTIC  
PROTEINS USING METADYNAMICS**

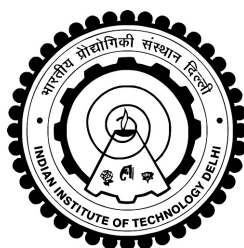
by

**Richa Singh**

Department of Chemical Engineering

*Submitted*

*in fulfilment of the requirements of the degree of Doctor of Philosophy  
to the*



**Indian Institute of Technology Delhi**

**October 2019**

*Dedicated to my family ...*

“The most beautiful thing we can experience is the mysterious. It is the source of all true art and science.”

-Albert Einstein

# Certificate

This is to certify that the thesis titled “**Thermodynamics, Kinetics, and Solution Structure Ensemble of Therapeutic Proteins Using Metadynamics**” being submitted by **Ms. Richa Singh** in the Department of Chemical Engineering, Indian Institute of Technology Delhi for the award of the degree of **Doctor of Philosophy**, is a record of bona-fide research work carried out by her under my guidance and supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The results contained in this thesis have not been submitted for the award of any other degree, associateship or similar title of any university or institution.

**Dr. Gaurav Goel**

Assistant Professor

Department of Chemical Engineering

Indian Institute of Technology Delhi

# Acknowledgements

I sincerely pay regards to my thesis adviser **Dr. Gaurav Goel** for his limitless patience and encouragement in guiding me through different learning curves of research. His idea of exploring the solution to a problem was never limited to trivial answers which constantly motivated me to go beyond my potential. The discussions I had with him were truly enlightening both in terms of understanding the subject as well as inculcating work ethics in me. His drive to solve the problems along, determination, proficiency, and insights have enabled me to work through even the most difficult challenges in my research. It is under his counsel that I have learned the role of perseverance, discipline and focus not only in achieving academic goals but also to move through the obstacles in life.

I also thank my research review committee members Prof. Anurag Singh Rathore, Prof. Sudip Pattanayek and Prof. Tapan Chaudhury for their timely and expedient inputs that really helped me to upgrade the quality of my research. The critical yet encouraging assessment of my work by the committee has guided me to fill in the gaps and improve the dearth of knowledge of the subject.

I especially want to acknowledge Parvez, my senior for being an inordinate source of support and help throughout. It is worth mentioning the names of my juniors- Rit Pratik, Julie and Akshay who never failed to assist me even with the smallest things I asked for. I also want to thank Nidhi, who has been a very good friend, a mentor and an important part of the memorable time I spent at IIT Delhi. A special mention to another friend and colleague Karan, for being extremely considerate in times of need. I am also thankful to my dearest friends- Hari, Namrata, Snigdha, Rajesh, Rami and Jawad who were always there to uplift my spirit even through the toughest times I had in my research.

I am grateful to the entire staff of the department of Chemical Engineering at IIT Delhi for their constant support at various levels. I am also thankful to IIT Delhi and the Department of Science and Technology for providing fellowship. Special acknowledgement to the central facility at IIT Delhi for dispensing computational

resources for my research and the entire HPC team for their regular and extremely valuable guidance on the operation of HPC.

Anything said would be less to praise my husband Ketan for his consistent support, strength and love. He stood by me through all the crests and troughs, constantly motivating me to work harder. I thoroughly admire him for being my buffer against stress throughout the time of research. I am truly indebted to my parents for their belief in me that culminated the necessary confidence to move forward and achieve my goals. No words can express my gratitude for them who provided me with infinite affection and liberty to choose. I also want to thank my brothers Avinash and Abhay and my sister-in-law Reeta for always standing with me as pillars of support and love. My journey at IIT Delhi would not have been easier without the bright moments I had shared with my lovely niece and nephew Arya and Shivansh. I have always strove to seek their delightful company for relieving myself of work pressure. I gratefully acknowledge my parents-in-law for their patience and faith that encouraged me to carry forward my academic interest. They definitely have a huge contribution to my achievements. I feel blessed to have as my partners in fun, my brother-in-law Chandan and his wife Aakriti. All our moments together while my short breaks from research hold a place very close to my heart. Credits to the little miracle, my daughter Ira who arrived into our lives timely enough, may be to get acknowledged in my thesis. She has been a wonderful child and her beautiful smile has always comforted me in the final days of my work.

This infinite world and its constituents, the balance of everything with its counterpart, like that of failure with success is not possible without the supervision of something that is beyond infinity. I thank that ineffable supervisor ‘God’ for all his favors and blessings in every part of my existence.

# Abstract

Knowledge of the solution ensemble of proteins, consisting of many short-lived intermediate conformations, can help in developing strategies to address two important protein-aggregation related challenges - (i) stability of liquid protein formulations and (ii) refolding yields of proteins. Experimental techniques used to characterize the solution ensemble provide important, but partial information as extensive time and ensemble averaging of data makes identification of individual short-lived intermediates difficult. Enhanced sampling molecular dynamics techniques, like metadynamics, have emerged as powerful computational tools that enable determination of the free energy landscape of protein folding as well as some other bio-molecular processes at a high spatial and temporal resolution. In metadynamics, the underlying hamiltonian is modified by adding a history-dependent potential that puts a penalty on already sampled regions of protein conformational space defined in terms of one or more collective variables (CVs), making possible the sampling of high free energy states. In this thesis, we have implemented well-tempered bias exchange metadynamics (WT-BEMD), a variant of metadynamics, to determine the solution structure ensemble and associated thermodynamics and kinetics of two important therapeutic proteins, namely insulin and transforming growth factor (TGF)- $\beta$ 3. Both proteins are known to aggregate under thermal and mechanical stress, thereby affecting their therapeutic applications. While sufficient spectroscopic and crystallographic information is available for insulin, TGF- $\beta$ 3 in turn, is a less investigated protein. Low solubility of TGF- $\beta$ 3 has long complicated its spectroscopic studies, although it is known that the active solution structure of TGF- $\beta$ 3 has several important differences with respect to its X-ray crystal structure.

WT-BEMD simulation of insulin was performed under high thermal and chemical stress replicating the amyloidogenic conditions used in several spectroscopic studies of this protein. A bin-based clustering method in-combination with a Markovian kinetic transition network model was employed on the BEMD trajectories to construct a thermodynamic and kinetic model of insulin folding. Major fraction ( 72 %) of the

obtained solution structure ensemble of insulin was observed to retain the native fold in form of the R-, R<sup>f</sup>- and T-states of this protein. A comparison of  $\alpha$ -proton chemical shifts of the obtained folded state ensemble (F) with those of the reported spectroscopic structure of insulin under amyloidogenic conditions provided an excellent correlation. The second largest fraction of structures corresponded to a compact denatured state ensemble (D) and the remaining structures belonged to an ensemble of partially folded states (PF) of insulin. Important differences in structure of the denatured and the partially folded intermediates were identified in terms of the local environments of amino acids. Moreover, in correspondence with the earlier findings, the breakdown of cooperativity in folding was observed for insulin where one of its partially folded states (folding intermediate) was found to connect the folded and the denatured states on the lowest free energy path (LFEP). These partially folded intermediates were compared and contrasted with spectroscopic, crystallographic, and calorimetric measurements during early stages of insulin aggregation. Further, a folding time of 0.46 ms for denatured to the folded state transition and an order of magnitude lower timescale for the partially folded to the folded state transition was identified using simulated annealing and kinetic Monte Carlo (KMC) algorithms.

WT-BEMD simulation of TGF- $\beta$ 3 was performed under typical re-folding conditions. Bin based clustering of the obtained BEMD trajectories provided four metastable state ensembles of this protein. These ensembles were exhaustively characterized based on several structural metrics (tertiary contact maps, residue based - solvent accessible surface area (SASA), root mean square deviation (RMSD),  $R_g$ , secondary structure content and  $\alpha$ -proton chemical shifts) and the nature of stabilizing interactions (hydrophobic, H-bonds, salt bridges, cation- $\pi$ ). A comparison of the entire solution ensemble constituted by these metastable state ensembles, with the reported spectroscopic structure of this protein provided excellent correspondence. Further, it was also found that the longest lived ensemble (constituting 50 % of the overall solution ensemble) provided the closest match with the reported details of solution structure of TGF- $\beta$ 3 and therefore, it was identified as the folded state of this protein in solution. This folded state was found connected to another long lived metastable ensemble containing denatured states, through a partially folded state on the lowest free energy path. Subsequently, the role of a zwitterionic detergent, CHAPS, in modulating the free energy landscape of TGF- $\beta$ 3 was studied by quantifying the preferential interaction coefficient of the obtained ensembles with CHAPS. The calculations revealed significant differences in solvation of native-like and partially folded / denatured states. Interestingly, two of these metastable ensembles showed signifi-

cant structural changes in short molecular dynamics simulations performed without CHAPS.

## सारांश

प्रोटीन के घोल की जानकारी, जिसमें कई अल्पकालिक मध्यवर्ती संरचनाएं शामिल हैं, प्रोटीन एकत्रीकरण से संबंधित दो मुख्य चुनौतियों का हल निकालने के लिए रणनीति बनाने में सहायता कर सकती है – (i) तरल प्रोटीन योगों की स्थिरता और (ii) प्रोटीन की कार्यपरक उत्पादन मात्रा। प्रोटीन के घोल का गुणात्मक परीक्षण करने के लिए उपयुक्त प्रायोगिक तकनीकें महत्वपूर्ण, किंतु आंशिक जानकारी प्रदान करती हैं क्योंकि इनसे उत्पन्न डेटा के व्यापक सामयिक और गुणात्मक औसत होने के कारण अल्पकालिक मध्यवर्ती प्रोटीन संरचनाओं की पहचान करना जटिल हो जाता है। इनहैस्ड सैंप्लिंग मोलिक्युलर डायनमिक्स तकनीकें, जैसे मेटाडायनमिक्स, शक्तिशाली कंप्यूटेशनल उपकरण के रूप में उभरी हैं जो प्रोटीन तह और साथ ही कुछ अन्य जैव आणविक प्रक्रियाओं के मुक्त उर्जा परिदृश्य को उच्च स्थानीय व सामयिक स्पष्टता से निर्धारण करने में सक्षम हैं। मेटाडायनमिक्स में, अंतर्निहित हैमिलटोनियन को इतिहास-निर्भर पोटेंशियल के संचय द्वारा संशोधित किया जाता है जिससे पहले से निर्धारित प्रोटीन संरचना क्षेत्र जो कलेक्टिव वेरिएबल्स (CVs) से परिभाषित होता है, के पुनः निर्धारण पर जुर्माना लगता है। इस थीसिस में, वेल-टेंपर्ड बायस एक्सचेंज मेटाडायनमिक्स (WT-BEMD) जो कि मेटाडायनमिक्स का एक प्रकार है, का उपयोग करके दो प्रमुख चिकित्साविधान प्रोटीनों, इंसुलिन और TGF- $\beta$ 3 की घोल में पाई जाने वाली संरचनाओं के समूह तथा उससे संबंधित उष्मागतिकी एवं बलगतिकी का निर्धारण किया गया है। दोनों प्रोटीन थर्मल और यांत्रिक तनाव से एकत्रित होने के लिए जाने जाते हैं, जिससे उनके चिकित्सकीय उपयोग पर असर पड़ता है। जहां इंसुलिन के लिए पर्याप्त स्पेक्ट्रोस्कोपिक एवं क्रिस्टलोग्राफिक जानकारी प्राप्य है, TGF- $\beta$ 3 इसके इतर, एक कम जांचा गया प्रोटीन है। TGF- $\beta$ 3 की कम घुलनशीलता ने लंबे समय से इसके स्पेक्ट्रोस्कोपिक अध्ययनों को जटिल बनाया है, हालांकि यह ज्ञात है कि TGF- $\beta$ 3 की घोल में पाई जाने वाली सक्रिय संरचना इसकी X-ray क्रिस्टल संरचना से काफी अलग है।

इंसुलिन का WT-BEMD सिमुलेशन उच्च तापीय और रासायनिक तनाव में किया गया था जो कि इस प्रोटीन के कई स्पेक्ट्रोस्कोपिक अध्ययनों में इस्तेमाल अमाइलॉइडोजेनिक स्थितियों की नकल है। इंसुलिन की तह के उष्मागतिकी एवं बलगतिकी मॉडल के निर्माण हेतु BEMD प्रक्षेपवक्रों पर बिन-आधारित समूहीकरण और मार्कोविअन काइनेटिक ट्रांजीशन नेटवर्क मॉडल का उपयोग किया गया था। प्राप्त इंसुलिन घोल संरचना समूह के मुख्य

अंश (७२ प्रतिशत) को इस प्रोटीन की R-, Rf- और T- अवस्थाओं में पाया गया है। प्राप्त की गई सक्रिय संरचना के समूह और इंसुलिन की अमाइलॉइडोजेनिक स्थितियों में कथित स्पेक्ट्रोस्कोपिक संरचना की  $\alpha$ -प्रोटॉन केमिकल शिफ्ट की तुलना करने पर उत्क्रिष्ट सहसंबंध मिलता है। संरचनाओं का दूसरा सबसे मुख्य अंश एक सघन विकृत संरचना समूह (D) के अनुरूप था और शेष संरचनाएं आंशिक रूप से बने संरचना (PF) समूह से संबंधित थीं। विकृत और आंशिक रूप से बनी मध्यवर्ती संरचनाओं में अमीनो एसिड के स्थानीय वातावरण के संदर्भ में महत्वपूर्ण अंतर पाया गया है। इसके अलावा, पहले के निष्कर्षों से मेल करता इंसुलिन तह प्रक्रिया में सहकारिता का टूटना अवलोकित किया गया है जहां इसकी एक आंशिक रूप से बनी अवस्था (मध्यवर्ती तह संरचना) को न्यूनतम मुक्त ऊर्जा पथ (LFEP) पर सक्रिय और विकृत संरचना के मध्य पाया गया है। इन आंशिक रूप से बनी मध्यवर्ती संरचनाओं की तुलना इंसुलिन एकत्रीकरण के शुरुआती चरणों में लिए गए स्पेक्ट्रोस्कोपिक, क्रिस्टलोग्राफिक और कैलोरीमीट्रिक मापों से की गई है। इसके अलावा, सिमुलेटेड अनीलिंग और काइनेटिक मॉटे कार्लो (KMC) कलन विधि के प्रयोग से विकृत से सक्रिय संरचना रूपांतरण का समय 0.46 ms और आंशिक रूप से बनी संरचना से सक्रिय संरचना रूपांतरण का समय एक परिमाण कम निर्धारित किया गया है।

TGF- $\beta$ 3 के WT-BEMD सिमुलेशन को प्रायः प्रयुक्त रीफोल्डिंग परिस्थितियों के अंतर्गत किया गया है। प्राप्त BEMD प्रक्षेपवर्कों के बिन-आधारित समूहीकरण ने इस प्रोटीन के चार मेटास्टेबल संरचना समूह प्रदान किये। इन समूहों का कई संरचनात्मक मापदंडों (टरशियरी संपर्क मानचित्र, रेसिड्यू-आधारित विलायक सुलभ क्षेत्र (SASA), रूट मीन स्क्वायर डेविएशन (RMSD),  $R_g$ , सेकेंडरी स्ट्रक्चर मात्रा और  $\alpha$ -प्रोटॉन केमिकल शिफ्ट) साथ ही स्थिरता प्रदान करने वाले संपर्कों (हाइड्रोफोबिक, H-बॉंड: साल्ट ब्रिज, कैटायन- $\pi$ ) की प्रकृति के आधार पर विस्तृत रूप से विश्लेषण किया गया। इन मेटास्टेबल संरचना समूहों से बने संपूर्ण घोल समूह की इस प्रोटीन की कथित स्पेक्ट्रोस्कोपिक संरचना से तुलना करने पर उत्क्रिष्ट समानता प्राप्त होती है। इसके अलावा, यह भी पाया गया है कि सबसे दीर्घ आयु वाला समूह (संपूर्ण घोल समूह का ५० प्रतिशत) TGF- $\beta$ 3 की कथित घोल संरचना के साथ सबसे निकटतम समानता दर्शाता है और इसीलिए, इसको इस प्रोटीन की घोल में सक्रिय संरचना माना गया है। इस सक्रिय संरचना को न्यूनतम मुक्त ऊर्जा पथ पर एक और दीर्घायु मेटास्टेबल समूह जिसमें विकृत संरचनायें हैं, के साथ एक आंशिक रूप से बनी संरचना के माध्यम से संपर्क में पाया गया है। तत्पश्चात, प्रिफ़रेंशियल इंटरैक्शन कोफ़ीशिएंट के मात्रात्मक ज्ञान से TGF- $\beta$ 3 के मुक्त ऊर्जा परिदृश्य को संशोधित करने में एक जुइटरायनिक डिटर्जेंट, CHAPS की भूमिका का अध्ययन किया गया है। गणना से पता चलता है कि सक्रिय समरूप संरचना और आंशिक रूप से बनी या विकृत संरचनाओं के विलायकयोजन में महत्वपूर्ण अंतर हैं। दिलचस्प बात यह है कि, इनमें से दो मेटास्टेबल समूहों ने CHAPS की अनुपस्थिति में की गई लघु आणविक गतिशीलता सिमुलेशन में महत्वपूर्ण संरचनात्मक बदलाव प्रदर्शित किये।

# Table of contents

|                                                              |           |
|--------------------------------------------------------------|-----------|
| Certificate . . . . .                                        | i         |
| Acknowledgements . . . . .                                   | ii        |
| Abstract . . . . .                                           | iv        |
| सारांश . . . . .                                             | vii       |
| List of figures . . . . .                                    | x         |
| List of tables . . . . .                                     | xii       |
| List of symbols . . . . .                                    | xv        |
| List of abbreviations . . . . .                              | 1         |
| <b>1 Introduction</b>                                        | <b>2</b>  |
| 1.1 Motivation . . . . .                                     | 2         |
| 1.2 Solution ensemble and folding of proteins . . . . .      | 3         |
| 1.3 Protein folding in experiments and simulations . . . . . | 7         |
| 1.4 Scope and objectives of thesis . . . . .                 | 9         |
| 1.5 Organization of thesis . . . . .                         | 10        |
| <b>2 Literature review</b>                                   | <b>13</b> |
| 2.1 Techniques to study protein folding . . . . .            | 13        |
| 2.2 Synergy between simulations and experiments . . . . .    | 22        |
| 2.3 Structure of partially folded state . . . . .            | 24        |
| <b>3 Methods and Materials</b>                               | <b>26</b> |
| 3.1 Metadynamics . . . . .                                   | 26        |
| 3.2 Simulation protocol . . . . .                            | 30        |
| 3.3 Post processing of the metadynamics data . . . . .       | 31        |
| 3.3.1 Thermodynamic model . . . . .                          | 31        |
| 3.3.2 Kinetic model . . . . .                                | 33        |
| 3.3.3 METAGUI generated microstate free energies for WT-BEMD | 34        |
| 3.3.4 Structural analysis of metastable ensembles . . . . .  | 35        |

|          |                                                                                                                                    |            |
|----------|------------------------------------------------------------------------------------------------------------------------------------|------------|
| 3.3.5    | NMR chemical shifts . . . . .                                                                                                      | 37         |
| 3.3.6    | Preferential interaction coefficient . . . . .                                                                                     | 38         |
| 3.3.7    | Lowest free energy path . . . . .                                                                                                  | 39         |
| <b>4</b> | <b>Thermodynamics, kinetics, and equilibrium ensembles of insulin in amyloidogenic conditions</b>                                  | <b>40</b>  |
| 4.1      | Introduction . . . . .                                                                                                             | 41         |
| 4.2      | Materials and methods . . . . .                                                                                                    | 44         |
| 4.3      | Results and discussion . . . . .                                                                                                   | 48         |
| 4.3.1    | Free Energy Landscape of insulin in CV space . . . . .                                                                             | 48         |
| 4.3.2    | Characteristics of Metastable States . . . . .                                                                                     | 53         |
| 4.3.3    | Lowest Free Energy Path from Folded to Denatured State . .                                                                         | 62         |
| 4.4      | Conclusion . . . . .                                                                                                               | 64         |
| <b>A</b> | <b>Appendices</b>                                                                                                                  | <b>66</b>  |
| <b>5</b> | <b>Solution structure ensemble of TGF-<math>\beta</math>3 and role of a zwitterionic detergent in modulating folding landscape</b> | <b>74</b>  |
| 5.1      | Introduction . . . . .                                                                                                             | 75         |
| 5.2      | Materials and methods . . . . .                                                                                                    | 78         |
| 5.3      | Results and discussion . . . . .                                                                                                   | 84         |
| 5.3.1    | Convergence of WT-BEMD simulation . . . . .                                                                                        | 84         |
| 5.3.2    | Solution ensemble of TGF- $\beta$ 3 . . . . .                                                                                      | 84         |
| 5.3.3    | Metastable ensembles of TGF- $\beta$ 3 in solution . . . . .                                                                       | 88         |
| 5.3.4    | Interaction with CHAPS and its effect on stability of metastable states . . . . .                                                  | 96         |
| 5.4      | Conclusion . . . . .                                                                                                               | 99         |
| <b>B</b> | <b>Appendices</b>                                                                                                                  | <b>101</b> |
| <b>6</b> | <b>Conclusion and future scope</b>                                                                                                 | <b>108</b> |
|          | <b>Bibliography</b>                                                                                                                | <b>112</b> |
|          | <b>Thesis based publications</b>                                                                                                   | <b>153</b> |
|          | <b>Author resume</b>                                                                                                               | <b>155</b> |

# List of figures

|     |                                                                                                                                                    |    |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 4.1 | Primary sequence of porcine insulin . . . . .                                                                                                      | 43 |
| 4.2 | Time evolution of 1D free energies along 7 PCs . . . . .                                                                                           | 49 |
| 4.3 | Average 1D metadynamics free energies plotted along first 5 PCs . .                                                                                | 51 |
| 4.4 | Metastable state hierarchy for insulin . . . . .                                                                                                   | 52 |
| 4.5 | C $\alpha$ probability contact map and $\alpha$ -proton chemical shifts for the folded ensemble . . . . .                                          | 54 |
| 4.6 | C $\alpha$ probability contact maps for the three partially folded states and the denatured state . . . . .                                        | 58 |
| 4.7 | Correlation plot for fraction of native hydrophobic contacts and relative exposed surface area for the ensemble of partially folded states . . . . | 60 |
| 4.8 | Property changes along the lowest free energy path connecting the central microstates of the folded state and the denatured state . . . . .        | 62 |
| A.1 | Spectral density plots of PC1, PC10, PC50 and PC153 . . . . .                                                                                      | 66 |
| A.2 | Plots showing diffusive dynamics of the 7 PCs obtained from regenerated WT-BEMD trajectories. . . . .                                              | 67 |
| A.3 | Average and difference contact maps of F, PF1-3 and D ensembles obtained from metadynamics . . . . .                                               | 68 |
| A.4 | Diagonal elements of diffusion rate matrix as a function of lag time obtained from simulated annealing maximization of the likelihood . .          | 68 |
| A.5 | Residue RMSF in F ensemble. . . . .                                                                                                                | 69 |
| A.6 | Residue Helicity in F, D, PF1, PF2, PF3 states and X-ray structure .                                                                               | 69 |
| A.7 | $A_{\text{rel}}$ for each residue in F, PF1-3 states and PF ensemble . . . . .                                                                     | 70 |
| A.8 | Fraction native hydrophobic contacts made by each residue in Denatured ensemble plotted as a function of $A_{\text{rel}}$ . . . . .                | 71 |
| 5.1 | Properties of CHAPS micelle . . . . .                                                                                                              | 80 |
| 5.2 | 1D free energies obtained after 1100 ns using sumhills utility of PLUMED-2.3 for CV1-7 . . . . .                                                   | 85 |

|      |                                                                                                                                          |     |
|------|------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.3  | Average 1D metadynamics free energies plotted along seven CVs for two halves of the converged trajectory . . . . .                       | 85  |
| 5.4  | RMSF, RMSD, C $\alpha$ -contact maps and $\alpha$ -proton chemical shifts for the obtained solution ensemble of TGF- $\beta$ 3 . . . . . | 87  |
| 5.5  | Probability C $\alpha$ contact maps of the metastable ensembles of TGF- $\beta$ 3 . . . . .                                              | 91  |
| 5.6  | Correlation plots between experimentally obtained and average $\alpha$ -proton chemical shifts for the metastable ensembles . . . . .    | 94  |
| 5.7  | Lowest free energy path of folding in 3D CV space plotted for TGF- $\beta$ 3 . . . . .                                                   | 96  |
| 5.8  | Local preferential interaction coefficient of the metastable ensembles of TGF- $\beta$ 3 with CHAPS . . . . .                            | 99  |
| B.1  | Sequence of Human TGF- $\beta$ 3. . . . .                                                                                                | 101 |
| B.2  | Range of CVs explored in preliminary BEMD and unbiased MD simulations . . . . .                                                          | 102 |
| B.3  | Transition time scales plotted using different sets of CVs using METAGUI . . . . .                                                       | 102 |
| B.4  | Average histograms and corresponding errors obtained from five converged WT-BEMD sub-trajectories plotted along CV(1-7). . . . .         | 103 |
| B.5  | Diffusive dynamics of CV(1-7) obtained for the converged WT-BEMD trajectory . . . . .                                                    | 103 |
| B.6  | Transition time scales after 1100 ns . . . . .                                                                                           | 104 |
| B.7  | Average C $\alpha$ contact maps of metastable ensembles and those obtained from their 200 ns unbiased MD simulations . . . . .           | 104 |
| B.8  | $A_{\text{rel}}$ of residues involved in salt bridges . . . . .                                                                          | 105 |
| B.9  | $\Gamma_{\text{XP}}^{i*}$ versus RMSF obtained from 10 ns unbiased MD simulations performed in absence of CHAPS . . . . .                | 105 |
| B.10 | Visual of CHAPS interaction with the most populated structure belonging to each metastable ensemble . . . . .                            | 106 |
| B.11 | RMSDs of metastable ensembles in 10 ns unbiased MD simulations in absence of CHAPS . . . . .                                             | 106 |

# List of tables

|     |                                                                                                                                                                                                                                                                                        |    |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.1 | Criteria for contacts . . . . .                                                                                                                                                                                                                                                        | 36 |
| 3.2 | Average $i$ , $i+3$ and $i$ , $i+4$ $C\alpha$ distance within helices in X-ray crystal structure of insulin. . . . .                                                                                                                                                                   | 37 |
| 4.1 | Total bias added in the first 150 ns and last 150 ns of the BE trajectory.                                                                                                                                                                                                             | 50 |
| 4.2 | Average properties of insulin metastable states classified in Figure 4.4.                                                                                                                                                                                                              | 53 |
| 4.3 | Distribution of contacts in metastable states. The contacts have been classified as native versus non-native as well as hydrophobic versus hydrogen bonds. In each case, a weighted-average value is determined from all the microstates belonging to a particular metastable state. . | 59 |
| A.1 | RMSDs from X-ray crystal structure and secondary structural elements in starting structures. . . . .                                                                                                                                                                                   | 67 |
| A.2 | MSD slopes for PCs, $D$ and PC widths. . . . .                                                                                                                                                                                                                                         | 71 |
| A.3 | Contribution of first 7 PCs from simulation s2 to express PCs obtained from simulation s1. . . . .                                                                                                                                                                                     | 72 |
| A.4 | RMSDs between central structure of F, PF1-3 and D states with analogous structural cluster in neutral replica. . . . .                                                                                                                                                                 | 72 |
| A.5 | Free energies of F, PF1-3 and D states in two halves after 150 ns. . .                                                                                                                                                                                                                 | 72 |
| A.6 | Populations of F, PF1-3 and D states in two halves after 150 ns. . . .                                                                                                                                                                                                                 | 73 |
| A.7 | Common contacts between metadynamics ensemble and its corresponding unbiased MD simulation trajectory. . . . .                                                                                                                                                                         | 73 |
| A.8 | Weighted average of the number of helical residues in F, D and PF1-3 states. . . . .                                                                                                                                                                                                   | 73 |
| 5.1 | List of intervals placed along each CV. . . . .                                                                                                                                                                                                                                        | 83 |
| 5.2 | Secondary structure <sup>a</sup> content in metastable states obtained from WT-BEMD . . . . .                                                                                                                                                                                          | 89 |

|     |                                                                                                                         |     |
|-----|-------------------------------------------------------------------------------------------------------------------------|-----|
| 5.3 | Comparison of stabilizing interactions in X-ray crystal structure and metastable states obtained from WT-BEMD . . . . . | 91  |
| 5.4 | Ensemble average properties of TGF- $\beta$ 3 metastable states classified in Figure 5.5. . . . .                       | 92  |
| B.1 | Hydrogen bonds with occupancies $>0.5$ . . . . .                                                                        | 101 |
| B.2 | Salt bridges. . . . .                                                                                                   | 107 |
| B.3 | Properties of microstates along LFEP . . . . .                                                                          | 107 |

# List of symbols

| Symbol            | Definition                                                             |
|-------------------|------------------------------------------------------------------------|
| $\omega$          | Gaussian potential deposition rate                                     |
| $\gamma$          | Bias factor                                                            |
| $\sigma$          | CV width                                                               |
| $\tau_G$          | Gaussian potential deposition time                                     |
| $\tau_b$          | Correlation time                                                       |
| $\Gamma_{XP}^*$   | Local preferential interaction coefficient as function of distance     |
| $\Gamma_{XP}^*$   | Global preferential interaction coefficient as function of distance    |
| $ds$              | Hypercube dimension                                                    |
| $s$               | CV value                                                               |
| $D$               | Diffusion coefficient                                                  |
| $F$               | Free energy                                                            |
| $f^S$             | 1D Free energy obtained from re-scaled Gaussian heights in HILLS files |
| $V_G$             | Bias potential                                                         |
| $W$               | Gaussian potential height                                              |
| $A_{rel}$         | Relative solvent exposed surface area                                  |
| $Q_{hydrophobic}$ | Fraction native hydrophobic contacts                                   |

# List of abbreviations

|         |                                          |
|---------|------------------------------------------|
| ANS     | 8-Anilinonaphthalene-1-sulfonic acid     |
| BEMD    | Bias exchange metadynamics               |
| CV      | Collective variable                      |
| FRET    | Fluorescence emission transfer           |
| FF      | Force field                              |
| HDX     | hydrogen/deuterium exchange              |
| HPC     | High performance computing               |
| MD      | Molecular dynamics                       |
| MC      | Monte-Carlo                              |
| MetaD   | Metadynamics                             |
| NMR     | Nuclear magnetic resonance               |
| PDB     | Protein data bank                        |
| PC      | Principal component                      |
| QM/MM   | Quantum Mechanics/ Molecular Mechanics   |
| RDC     | Residual dipolar coupling                |
| REMD    | Replica exchange metadynamics            |
| RMSD    | Root mean square deviation               |
| RMSF    | Root mean square fluctuation             |
| SASA    | Solvent Accessible Surface Area          |
| SAXS    | Small Angle X-ray Scattering             |
| SMD     | Steered Molecular Dynamics               |
| UV CD   | Ultra Violet Circular Dichroism          |
| WT-BEMD | Well Tempered Bias Exchange Metadynamics |