# COMPUTATIONAL OPTIMIZATION ALGORITHMS FOR PROTEIN STRUCTURE REFINEMENT

---

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

---

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

DEBSWAPNA BHATTACHARYA

Professor Jianlin Cheng, Thesis Supervisor

DECEMBER 2014

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled:

COMPUTATIONAL OPTIMIZATION ALGORITHMS
FOR PROTEIN STRUCTURE REFINEMENT

presented by Debswapna Bhattacharya,

a candidate for the degree of master of science and hereby certify that, in their opinion, it is worthy of acceptance.

---

Professor Jianlin Cheng

---

Professor Dong Xu

---

Professor Xiaoqin Zou

# DEDICATION

To my parents

for inspiring me to follow my dream

# ACKNOWLEDGMENTS

I would like to thank my advisor, Professor Jianlin Cheng, for his patient advice, guidance and inspiration over the years and giving me so much freedom and opportunity to explore different aspects of computational structural biology. I am greatly indebted to Professor Cheng for all that I learnt from him both in terms of technical knowledge and research style, and this work would have been impossible without his support. My other committee members have also been very supportive. Professor Dong Xu has long been an inspiration to me. His classes and presentations have proved to be one of my best learning experiences. Professor Xiaoqin Zou proved to be a most through reviewer of my work and consistently provided honest and constructive feedback.

I would like to thank my colleagues and coworkers proving me a brilliant, cooperative environment and for many enlightening and edifying discussions and collaborations. These include Zheng Wang, Jesse Eickholt, Xin Deng, Jilong Li, Renzhi Cao, Badri Adhikari, Qi Qi and all the members of the Bioinformatics, Data Mining, Machine Learning Laboratory.

I would like to thank Dr. Chen Keasar of Ben-Gurion University, Israel, for making the MESHI molecular modeling suite available, which serves as the framework for implementing the protein structure refinement methods, 3Drefine and i3Drefine. I would also like to thank numerous users of 3Drefine and i3Drefine who have contributed bug fixes and improvements.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

**CM**: Comparative Modeling

**PDB**: Protein Data Bank

**NNSM**: Near Native Structure Model

**NS**: Native Structure

**MM**: Molecular Mechanics

**MD**: Molecular Dynamics

**TBM**: Template Based Modeling

**FM**: Free Modeling

**TS**: Tertiary Structure

**SC**: Side Chain

**CASP**: Critical Assessment of protein Structure Prediction

**3D**: Three-Dimensional

**HB**: Hydrogen Bonding

**RMSD**: Root Mean Square Deviation

**GDT**: Global Distance Test

**TM**: Template Modeling

**CAD**: Contact Area Difference

**L-BFGS**: Limited memory Broyden–Fletcher–Goldfarb–Shannon

**PR**: Problematic Region

# ABSTRACT

Despite significant advancement being made during the recent past in predicting structure of proteins using computational methods, these techniques often cannot achieve sufficiently high level of accuracy to fully appreciate biological function and to serve as a reliable starting point for rational drug design efforts to develop novel therapeutics. Bringing these low-resolution models as close as possible to the native structure, called the protein structure refinement problem, however, has remained largely unsolved. Existing approaches in protein structure refinement suffer from two key challenges: (1) lack of consistency and (2) failure to produce meaningful degree of refinement.

This thesis is composed of three major contributions. First, we propose a consistent and computationally efficient computational optimization protocol called 3Drefine. Next, we further improve the 3Drefine algorithm by developing an iterative version of the protocol, named i3Drefine. Finally, we present a novel conformation ensemble-based iterative refinement method, REFINEpro, aimed at producing pronounced degree of refinement. All of these methods were benchmarked in large-scale benchmark datasets and achieved consistent refinement in both global and local structural quality measures. In particular, i3Drefine was ranked as the best protein structure refinement server method in recent Critical Assessment of Protein Structure Prediction experiment. All of these methods are freely available to the scientific community in the form of software and web-servers.

# Chapter 1

# Introduction

In this chapter, we give an overview of protein structure refinement problem that we will consider in this thesis. We also describe some of the issues faced in computational approaches for protein structure refinement that we will try to address, and give an outline for rest of this thesis.

## 1.1 Introduction to Protein Structure Refinement

Substantial progress has been made during the recent past toward the longstanding goal of computationally predicting the structure of proteins from their amino acid sequence [1-4]. Specifically, using templates from known structures of homologous proteins and advanced sequence-template alignment techniques (a.k.a. comparative modeling; CM) [5, 6], generation of reliable models capturing at least the overall fold topology of unknown protein structures is made possible. Moreover, the steady growth of available experimental protein structures in the Protein Data Bank (PDB) [7], especially from structural genomics initiatives [8-11], has led to a near complete coverage of protein fold

space [12, 13]. Consequently, it is possible to predict the structure of most unknown proteins to some accuracy through comparative modeling [14, 15]. In these types of modeling applications, the predicted structure is typically within 3Å Cα root mean square deviation (RMSD) range of the true experimental structure (a.k.a. native structure) [16]. These predicted models are sometimes also referred to as near-native structure models (NNSM) due to their close proximity to native structure (NS). An example to NNSM and NS of an example protein (4pwu) has been shown in Figure 1.1. Unfortunately, NNSM is not at sufficiently high level of accuracy to fully elucidate biological function and to serve as a reliable starting point for rational drug design efforts to develop novel therapeutics [17, 18]. Bringing the NNSM as close as possible to the NS to facilitate biological relevance is known as protein structure refinement problem.

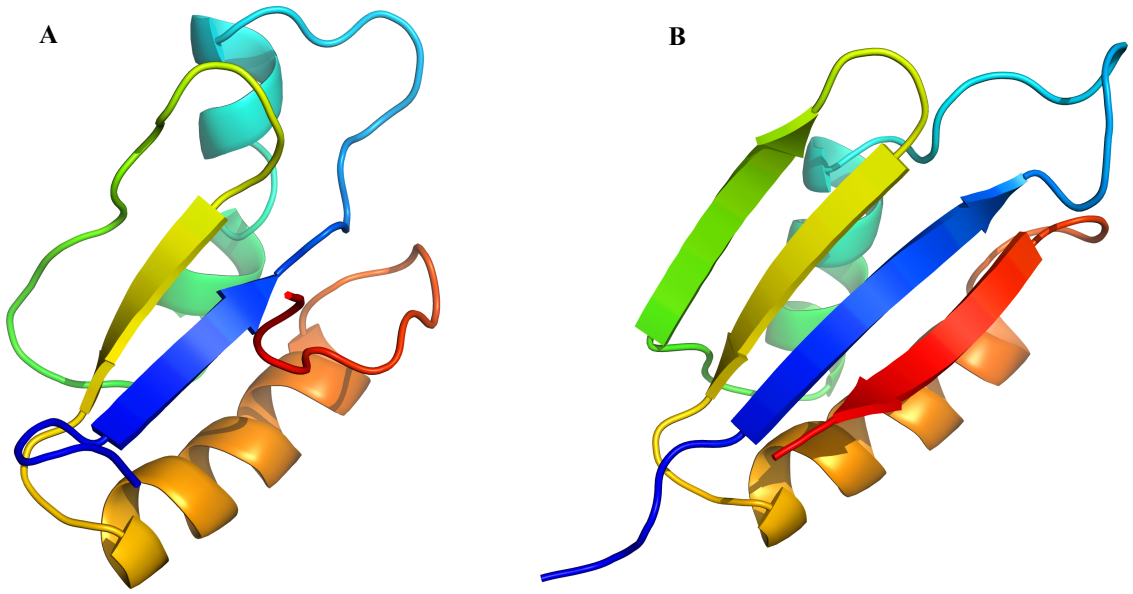


**Figure 1.1**. Near Native Structure vs. Native Structure of an example protein (4pwu).
(A) Near Native Structure. The structure has been rainbow colored from N-terminus to C-terminus.
(B) Native Structure. The structure has been rainbow colored from N-terminus to C-terminus.

## 1.2 Existing Approaches and Challenges in Protein Structure Refinement

Efforts to solve the protein structure refinement problem have usually been rooted in two schools of thoughts. One is physics based methods which is governed by the thermodynamic hypothesis proposed by Anfinsen that the native structure of a protein corresponds to the global minimum of its free energy [19]. Consequently, a force-field is first developed to calculate the potential energy of the initial protein model. Then the potential energy is minimized through conformation changes with the goal to find the free-energy minimum in the protein energy landscape using traditional molecular mechanics (MM) potentials like AMBER99 [20, 21], OPLS-AA [22], etc. A number of noteworthy studies have been performed in this direction [23, 24]. However, there are two major bottlenecks of these methods: (1) limited accuracy of physics based empirical force fields and (2) "multiple-minima problem" arising from the presence of many local minima in protein's multidimensional energy landscape [25]. The other school of thoughts is "knowledge-based" methods that utilize the statistical potentials derived from the analysis of recurrent patterns in experimentally derived protein structures and sequences [26]. Molecular Dynamics (MD) simulation is widely used in this kind of protocols [27, 28] to move every atom of a protein. Apart for some isolated cases, however, no systematic structural improvement has been attained [29].

Some promising progress has been made in the recent past by combining the two schools of thoughts, that is, by using composite physics and knowledge-based potentials [30-32] to solve protein structure refinement problem. Although encouraging, these techniques

highlight two key issues in protein structure refinement – (1) lack of consistency and (2) failure to produce meaningful degree of refinement. That is, majority of these methods either follow a computationally expensive robust sampling strategy around the starting structure in order to produce large degree of refinement at the cost of consistency, or a conservative local sampling producing improvement only in local qualities of the models rather than substantially improving the backbone positioning. Development of methods capable of tacking these two problems is, therefore, a crucial step forward for solving protein structure refinement problem and more generally, towards the improvement of computational protein structure prediction.

## 1.3 Community-wide Assessment of the Progress in Protein Structure Refinement

The biennial community-wide Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiment aims to evaluate the progress and challenges in the state-of-the-art of protein structure modeling techniques, one of the fundamental problems in computational biology - prediction of the tertiary structure of protein from its sequence information. During the recent CASP experiments, encouraging and consistent progress have witnessed in template-based modeling (TBM) [33-36] or ab-initio (free-modeling; FM) [37-40] folding of protein structures. The refinement category has been a recent addition to the CASP framework since CASP8, which aims to evaluate whether further improvement is possible to the best predictions made by contemporary structure

prediction techniques. In the blind refinement experiment, predictors are given a starting structure evaluated by the organizers as the best submitted model during the structure prediction phase (TS category) along with the sequence information. Occasionally, some hints are also provided to aid the refinement like the focus regions during refinement or the accuracy of the starting structure.

Since its inclusion during CASP8, refinement category has been drawing increasing attention by the community. During recently concluded CASP10 refinement experiment, a 92% increase in the number of refinement targets and 39% increase in the number of participating groups have been observed compared to CASP9. This is not unexpected because a consistent and efficient refinement protocol can serve as a natural end step in almost all the contemporary structure prediction pipelines adding value to the already predicted structures through simultaneous improvement in backbone geometry and correction of local errors like irregular hydrogen bonding, steric clashes, unphysical bond length, unrealistic bond angles, torsion angles and side-chain $\chi$ angles. However, structure refinement has proven to extremely challenging as revealed in the assessment of refinement experiments during CASP8 and CASP9 [41, 42] with only a few participating groups were able to improve the model quality consistently. It should be noted, however, that CASP refinement category differs in a slight but significant way from refinement in the context of TBM [23, 27, 43-49] where the objective is to refine the best identified template structure(s) to produce better quality prediction. In CASP, on the other hand, the starting models issued for refinement have already been refined by other structure prediction pipelines and judged to be the best among all the submitted models. Thus,

attempts to improve qualities of these models would naturally impose more challenges and often the risk of degrading the model quality instead of improving it.

## 1.4 Thesis Outline and Contributions

The remainder of this thesis is structured as follows. In chapter 2, we begin be attempting to address the first challenge associated with protein structure refinement – consistency. We propose a two-step refinement protocol, called 3Drefine, to consistently bring the initial model closer to the native structure. The first step is based on optimization of hydrogen bonding (HB) network and the second step applies atomic-level energy minimization on the optimized model using a composite physics and knowledge-based force fields. When evaluated on the CASP benchmark data, we found that around 80% of the times 3Drefine has improved the global qualities in the starting structures in CASP8 and CASP9 refinement targets. Also, the ability of 3Drefine to simultaneously improve the backbone positioning and local model qualities is encouraging. The contents of Chapter 2 and Appendix A are mostly from the manuscript published as:

*Bhattacharya, D. and Cheng, J. (2013), 3Drefine: Consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization.* ***Proteins: Structure, Function, and Bioinformatics****, 81: 119–131. doi: 10.1002/prot.24167.*

In Chapter 3, we further improve the 3Drefine algorithm by developing an iterative version of the protocol, named i3Drefine and participated in CASP10 refinement category in order to test the method in blind mode. As per the official CASP10 results

released during CASP10 meeting in the form of assessors' presentation (http://predictioncenter.org/casp10/docs.cgi?view=presentations), i3Drefine was ranked as the single best refinement server method capable of consistent improvement in qualities of starting structures. The contents of Chapter 3 is mostly from the manuscript published as:

*Bhattacharya, D. and Cheng, J. (2013), i3Drefine Software for Protein 3D Structure Refinement and Its Assessment in CASP10.* ***PLoS ONE*** *8(7): e69648. doi:10.1371/journal.pone.0069648.*

In Chapter 4, we turn our attention to address the second challenge related to protein structure refinement – degree of refinement. We presented a new conformation ensemble-based iterative refinement method, REFINEpro, aimed at substantially improving the overall fold of the initial models through refinement of loop and terminal regions or rearrangements of disoriented secondary structure segments, accompanied by correction of local errors. By performing a large-scale benchmark study on 163 targets, we demonstrated that the protocol is capable of pronounced improvement in global and local qualities of protein models generated by both TBM and FM methods. The contents of Chapter 4 is mostly from the Association for Computing Machinery (ACM) conference proceeding published as:

*Bhattacharya, D. and Cheng, J. (2013), Protein Structure Refinement by Iterative Fragment Exchange.* ***Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics****. ISBN: 978-1-4503-2434-2, doi: 10.1145/2506583.2506601****.***

Finally, in Appendix B, we provide a brief overview of the freely available software and web-services developed based on the aforementioned methods and protocols for the scientific community. These freely available software and web servers would allow researchers from around the world to apply these methods to their own data and these fully automated and computationally inexpensive systems provide a suitable framework for high-throughput proteomics and protein engineering projects.

To summarize, the contributions of this thesis are two-fold: (1) attempting to address two key issues in existing approaches of computational protein structure refinement – lack of consistency and failure to produce meaningful degree of refinement and (2) providing the scientific community with access to fast, reliable and freely available refinement software and web-services to facilitate biomedical research.

# Chapter 2

# 3Drefine: Consistent Protein Structure Refinement by Optimizing Hydrogen Bonding Network and Atomic-Level Energy Minimization

## 2.1 Abstract

One of the major limitations of computational protein structure prediction is the deviation of predicted models from their experimentally derived true, native structures. The limitations often hinder the possibility of applying computational protein structure prediction methods in biochemical assignment and drug design that are very sensitive to structural details. Refinement of these low-resolution predicted models to high-resolution structures close to the native state, however, has proven to be extremely challenging. Thus, protein structure refinement remains a largely unsolved problem. Critical assessment of techniques for protein structure prediction (CASP) specifically indicated that most predictors participating in the refinement category still did not consistently improve model quality.

Here, we propose a two-step refinement protocol, called 3Drefine, to consistently bring the initial model closer to the native structure. The first step is based on optimization of hydrogen bonding (HB) network and the second step applies atomic-level energy minimization on the optimized model using a composite physics and knowledge-based force fields. The approach has been evaluated on the CASP benchmark data and it exhibits consistent improvement over the initial structure in both global and local structural quality measures. 3Drefine method is also computationally inexpensive, consuming only few minutes of CPU time to refine a protein of typical length (300 residues).

3Drefine web server is freely available at http://sysbio.rnet.missouri.edu/3Drefine/.

## 2.2 Introduction

The goal of protein tertiary structure prediction is to accurately estimate the three-dimensional position of each atom in a protein. Comparative modeling (or homology modeling) is the most widely used technique in the field of protein structure prediction. In the traditional comparative modeling methods, an experimental protein structure (or template) that has significant sequence similarity to the target protein of interest is first identified, and then a sequence alignment between the target and the template is generated in order to use the structural information of the aligned regions of the template to construct a structural model for the target protein[50]. But, even with the best possible template and target sequence alignment, predicted models often deviate from the true native structures in terms of their atomic coordinates. Significant progress toward

improving the accuracy of comparative modeling has been made during recent years by building the target structure combining the fragments from multiple templates [51-53]. The introduction of multiple templates has certainly enhanced the accuracy of structure prediction by bringing the model closer to the native structure than using a single template. Despite having largely correct backbone conformations, these models sometimes still have poor structural qualities, including irregular hydrogen bonding network, steric clashes, unphysical bond length, unrealistic bond angles, torsion angles and side-chain $\chi$ angles. Thus, direct refinement of the predicted models from their coordinate information alone with the goal of detection and correcting the errors is an essential part of computational protein structure prediction process.

The earlier studies in the field of protein structure refinement can be broadly classified into two categories: (1) methods that perform significant conformational changes in terms of backbone positioning and structural information [54-56] and (2) protocols that make small structural changes at the local level by modifying the side chain conformation[57, 58] or removing the gaps and steric clashes [59, 60]. The first kind of methods is more adventurous having potential of substantial improvement in the structural qualities [61]. But these techniques are computationally expensive and often inconsistent. The second type of protocols, although consistent, generally fails to bring model considerably closer to the native state. Consistent and simultaneous improvement in both global and local structural qualities of the initial models in a computationally efficient manner is, therefore, a nontrivial problem.

Protein structure refinement has received noteworthy attention in the recent critical assessment of techniques for protein structure prediction (CASP8 and CASP9).

According to the results in CASP9 refinement experiment, very few methods exist which can consistently bring the initial models closer to the native structure. The majority of the improvement has been observed at the local level by modification of the physicality of the models or alteration of the side chain positions and not at the global topologies [41, 62]. Encouragingly, some promising progress has been made in the recent past by development of methods that have the potential to improve both the global topologies and the local structural qualities of the predicted models by using optimized physics-based all-atom force field [63], applying knowledge-based potential [30, 64] or performing Fragment-Guided Molecular Dynamics Conformation Sampling [31]. These protocols usher the way to solve the protein structure refinement problem.

In this work we present an efficient refinement protocol, called 3Drefine that is based on two steps of refinement process. We extensively test this method on CASP benchmarks having high diversity in the difficulty of the prediction targets. 3Drefine demonstrates significant potential in atomic-level protein structure refinements in terms of both global and local measures of structural qualities. Thus, we expect the protocol to be a useful addition to current state-of the-art refinement tools. We also hope that this method can be adopted as a final step in the existing protein structure prediction pipeline.

## 2.3 Materials And Methods

The 3Drefine protocol refines the initial protein structures in two steps: (1) Optimizing Hydrogen Bonds network and (2) atomic-level energy minimization using a combination of physics and knowledge based force fields; both carried out using Java based molecular

modeling package MESHI[65]. The justification for choosing MESHI over the other modeling packages is threefold: (1) the strict object-oriented design used in MESHI enhances possibility of code reuse by means of inheritance mechanism provided in Java, thereby reducing the development time; (2) MESHI has robust garbage-collection utility to deal with failures and (3) the open source and platform independent nature of MESHI makes it more flexible.

### 2.3.1 Optimizing Hydrogen Bonds Network

Hydrogen atoms are the most frequently occurring atoms in a protein structure and play a crucial role in protein folding through hydrophobic interaction or hydrogen bonding[66-68]. Previous studies in the field suggest that accurately determining the positions of the hydrogen atoms has a major influence on applying atomic-level potentials for protein structure refinement[69-71]. Unfortunately, most of the computational methods for protein structure prediction lack the ability to consistently and correctly identify the hydrogen atoms. We, therefore, decided to first optimize the hydrogen bonding network in the initial model.

The current state-of-the-art protocols for predicting the hydrogen bonds generally follow a combination of local geometry restraint and a conformational search [72-75]. We adopt a very similar approach here to optimize hydrogen bonding network. Using MESHI [65], the hydrogen bonds in the initial full-atomic model are calculated first. The position of non-polar hydrogen atoms are determined by using fixed bond lengths and bond angles parameters supplied in the MESHI library which have been derived by a collection of

1145 protein domains as part of MESHI package. For the polar hydrogen atoms, a search is performed via “Geometry” and “Molecular Elements package” of MESHI to find out the most favorable position of hydrogen atoms satisfying hydrogen bonds with the closest neighboring atoms and considering the protonation state of each amino acid. We call this an extended atomic model.

### 2.3.2 Atomic-level Energy Minimization

We use atomic-level force fields driven by MESHI [65] for performing energy minimization on the extended atomic model. Since the current release of MESHI package does not include any established force fields, we, therefore, construct a customized all-atomic total energy of a protein model by combining the energetic contributions of the bonded interactions described in ENCAD potential [76], which is an example of a traditional molecular mechanics force field, some standard energy elements using the “Energy package” included in MESHI software and a Knowledge-based atomic pairwise potential of mean force [64]. The ENCAD molecular mechanics force fields are chosen because they are freely available and have all been implemented for use with the MESHI molecular dynamics package. We include only the bonded terms of the ENCAD potential (bond stretches, bond angle bends, and torsion angle twists) since they are stronger than the nonbonded terms [64] (van der Waals interactions, and electrostatic interactions). As some of the energy terms require the secondary structure values for accurate calculations, we set the secondary structure of the extended atomic model by using DSSP [77]. Almost all energy term requires some knowledge about the distances between the atoms. To this

end, we use a fast heuristics for calculating distances in the system. Given the atom list of the system, an internal matrix of distance objects is created for all the inter-atomic distances by means of the "Distance Matrix class" in MESHI. Following the standard convention adopted in the MESHI package to calculate distances in a computationally inexpensive way, we have made two assumptions: first, the atoms that are separated by 4 bonds or less are considered bonded and second, distances must be within a cutoff of 5.5 Å. This means that any distance between non-bonded atoms (separation of more than four covalent bonds) that is higher than 5.5 Å has been assumed infinite.

Finally, the customized total energy of the extended atomic model, which is used to guide the minimization, is calculated by MESHI "Total Energy class" and consists of the following terms:

$$E_{total} = E_{bondlength} + E_{bondangle} + E_{torsion} + E_{hydrogenbonds} + E_{tether} + E_{kbpairwise} \quad (1)$$

In the following sections we will describe each of the energy terms mentioned in Eq. (1):

### 2.3.2.1 Bond Length Energy Term

When two atoms are connected by a chemical bond they tend to maintain a fixed distance (called equilibrium distance) depending on the type of the atoms participating in the bond formation. Any change in this equilibrium distance adds potential energy to the protein. As per the ENCAD potential [76], this is usually represented as:

$$E_{bondlength} = \sum \frac{1}{2} K_b (b - b_0)^2 \quad (2)$$

where b is the distance between the two bonded atoms, $b_0$ is their equilibrium distance and $K_b$ is a bond stretching force-constant subject to the atom types. The parameters $b_0$ and $K_b$ depend on the type of the bonds and their values can be found at the published work for ENCAD bonded energy terms [76]. This is denoted in Eq. (1) as $E_{bondlength}$ term.

### 2.3.2.2 Bond Angle Energy Term

Similar to bond length, when three atoms are connected with two chemical bonds they incline to maintain a fixed angle (called equilibrium angle) subject to the atom types. Any variation in this equilibrium angle adds potential energy to the protein. Following the standard ENCAD potential [76], this can be defined as:

$$E_{bondangle} = \sum \frac{1}{2} K_\theta (\theta - \theta_0)^2 \quad (3)$$

where θ is the distance between the two bonded atoms, $\theta_0$ is their equilibrium distance and $K_\theta$ is a bond angle force constant. Like bond length energy term, the parameters $\theta_0$ and $K_\theta$ related to bond angle energy term depend on the type of the bonds the atoms are involved in and their values can be found at the publication of the ENCAD bonded energy terms [76].We symbolize this in Eq. (1) as $E_{bondangle}$ term.

### 2.3.2.3 Dihedral Or Torsions Angle Twist Energy Term

The third term in Eq. (1) has been denoted as $E_{torsion}$ and it represented in ENCAD potential [76] as:

$$E_{torsion} = \sum \frac{1}{2} K_\emptyset [\, 1 - \cos n(\emptyset_i - \emptyset_0)] \quad (4)$$

where n is the periodicity, $\varphi_0$ is the equilibrium value and $K_\varphi$ is the half of the rotation barrier height. Values of these parameters have been described in ENCAD potential [76]. The torsion energy term has the ability to represent true dihedral angles and unrealistic or out-of-plane torsion (or dihedral) angles.

We implement all the above mentioned energy terms of the bonded interactions using the MESHI framework [65, 78].

### 2.3.2.4 Hydrogen Bonds Energy Term

Hydrogen Bonds energy term calculates the energy over all the backbone Hydrogen bonds in the protein and is denoted by $E_{hydrogenbonds}$ in Eq. (1). We use a combination of *"Energy"* and "*Geometry packages*" of MESHI [65] framework to calculate the energy of the hydrogen bonds for the extended atomic model. Following the explicit hydrogen bonding potential defined in FG-MD [31] refinement method, we consider only the short range hydrogen-bonding potential with cutoff distance between the Hydrogen and the Oxygen $\leq$ 3 Å. This is defined as:

$$E_{HB}(d_{ij},\alpha,\beta) = \begin{cases} k_1(d_{ij} - d_0)^2 + k_2(\alpha - \alpha_0)^2 + k_3(\beta - \beta_0)^2 & d_{ij} \leq 3.0 \\ 0 & d_{ij} > 3.0 \end{cases} \quad (5)$$

where $d_{ij}$ is the distance between hydrogen of the donor and oxygen of the acceptor, α is the N-H-O angle and β is the angle of C-O-H. Values of these parameters have been adopted from the published work of FG-MD[31] as $d_0$ = 1.95 $\pm$ 0.17 Å, $\alpha_0$ = 160.0 $\pm$ 12.2°, $\beta_0$ = 150.0 $\pm$ 17.5° and the values of the force constants are $k_1$ = 2.0, $k_2$ = 0.5 and $k_3$ = 0.5.

### 2.3.2.5 Tethering Energy Term

Protein models sometimes have unfavorable atomic interaction and these disordered atomic positions can cause large initial forces that result in artificial movement away from the original structure while performing energy minimization. One solution to avoid these large deviations is to relax the protein models gradually. But a more profound approach would be to assign tethering forces to all heavy atoms during the minimization process. The tethering constant is a force applied to fix atomic coordinates on predefined positions and the strength of tethering force affects the extent of movement of the atoms from the initial coordinates. While tethering the well-defined main chain atoms, the side chains are allowed to move and adjust their position in order to minimize the total potential energy. Tethering of protein is known to have significant impact on the rates and mechanisms of protein folding [79]. Tethering energy term, symbolized by $E_{tether}$ in Eq. (1) is a tethering term of the $C_\alpha$ and $C_\beta$ atoms of the model to their initial positions. We implemented the tethering energy term by means of the "*Molecular Elements package*" in the MESHI software. Tethering spring constants have been set to 1 Energy Unit/Å.

### 2.3.2.6 Knowledge-Based Atomic Pairwise Potential Of Mean Force Energy Term

The final term in Eq. (1), $E_{kbpairwise}$ is an implementation of the knowledge-based potential of mean force [64]. The original work is based on the interaction statistics of 167 atom types derived by counting of pairwise atomic contact frequencies of proteins from a selection of 500 files from the Protein Data Bank (PDB) having high resolution (1.8 Å or

better), low homology, and high quality. In the original study, the weight (w) of KB01 [64] potential has been set to w = 1.0, which is near-optimally weighted [30]. We use the same weight for our refinement protocol. The knowledge-based potential of mean force has been implemented via "*Energy package*" of MESHI [65].

### 2.3.3 Minimization Protocol

The 3Drefine minimization involves 200,000 steps of energy minimization using limited memory Broyden-Fletcher-Goldfarb-Shannon (L-BFGS) algorithm [80] or until convergence to machine precision, which is carried out by the "*Optimizers package*" in MESHI [65] framework. The backbone structure is refined primarily by the bonded terms of the ENCAD potential [76] and the knowledge-based potential of mean force energy term [64] while the Tethering energy term plays crucial role in optimizing the side chains. The hydrogen bonding network is updated during the minimization by using the explicit Hydrogen bonding energy term described in Eq. (5). The energy minimized model is the final refined model.

### 2.3.4 Data Set Preparation

To assess the performance of 3Drefine approach, we collected the refinement targets on recent critical assessment of techniques for protein structure prediction (CASP) [81]. To further test the protocol on a large benchmark of 107 CASP9 targets, we used the initial models generated by our structure prediction method, MULTICOM-REFINE [82],

participated during the CASP9 experiment. The structure refinement category has been introduced since CASP8 [81]. In these experiments, the predictors were given a starting model for refinement in a blind mode. These starting models had been generated by the CASP structure prediction servers and the organizers evaluated it to be among the best prediction for each target. Although 3Drefine run has been performed after the CASP8 and CASP9 refinement experiments, we ensured same modeling conditions as the CASP blind predictors so that the performance of 3Drefine can be directly compared with the other state-of-the-art refinement methods participating in CASP8 and CASP9.

## 2.3.5 Metrics Used For Evaluation

We determine the quality of the structural refinement by observing the changes in global topologies of the models before and after refinement with respect to their native structures. We also determine the local structural qualities of the initial and refined model in order to measure the physical reasonableness of the structure. We have focused on GDT-HA score [83], TM-score [84] and $C_\alpha$ RMSD [85] which are measures of the global positioning of $C_\alpha$ atoms. To evaluate the local qualities of the models, we use MolProbity score [86].

### 2.3.5.1 GDT-HA Score

GDT-HA [83] score measure the fraction of Cα atoms that are positioned correctly. It counts the average percentage of residues with $C_\alpha$ distance from the native structure residues below 0.5, 1, 2, and 4 Å, respectively, after optimal structure superposition. GDT-HA is related to GDT-TS, which uses cutoffs of 1, 2, 4, and 8 Å. Therefore, GDT-

HA is more sensitive to small structural errors. Because of the strong mutual correlation between GDT-HA and GDT-TS score, we chose to only use GDT-HA in our analysis. GDT-HA score has been a widely used scoring function to measure the global positing of $C_\alpha$ atoms in CASP experiments [33, 36, 87]. It ranges from [0, 1] with higher value indicating better accuracy.

### 2.3.5.2 Template Modeling Score (TM-Score)

TM-score [84] is a variation of the Levitt–Gerstein (LG) [88] score. It is a global measure of similarity of structural topologies. TM-score is defined as follows:

$$\text{TM} - \text{score} = max\left[\frac{1}{L_N}\sum_{i=1}^{L_T}\frac{1}{1+(d_i/d_0)^2}\right] \quad (6)$$

where $L_N$ is the length of the native structure, $L_T$ is the length of the aligned residues to the template structure, $d_i$ is the distance between the $i^{th}$ Cα pair of aligned residues, and $d_0$ is a scale to normalize the match difference. "*max*" represents that the maximum value is considered after optimal spatial superposition. The value of the constant $d_0$ is expressed in the original work [84] as:

$$d_0 = 1.24\sqrt[3]{L_N - 15} - 1.8 \quad (7)$$

Like GDT-HA score, TM-score also lie in [0, 1] with higher TM-score suggests enhanced accuracy. However, rather than using specific distance cutoffs and focusing only on the fractions of structures as described for GDT-HA score, all the residues of the modeled proteins are evaluated in the TM-score. Furthermore, TM-score does not depend

on the protein length. A TM-score > 0.5 indicates that the proteins share the same fold [89].

### 2.3.5.3 RMSD

We evaluated the Root Mean Square Deviation (RMSD) [85] of the $C_\alpha$ positions of the atoms in order to determine the average distance between the $C_\alpha$ atoms after superposition. Similar to GDT-HA and TM-score, RMSD is a global measure of the correct positioning of the $C_\alpha$ atoms. However, since RMSD is based on a single superposition, lacking any kind of distance cutoffs, there is a week correlation between GDT-HA and RMSD score. Also, unlike GDT-HA or TM-score; a lower RMSD value indicates that the protein model is close to its native state. Even if the coordinates of only a few atoms undergo large atomic changes, RMDS becomes high; making RMDS sensitive to small structural errors.

### 2.3.5.4 MolProbity

MolProbity [86] is an all-atom measure of the physical correctness of a structure based on statistical analysis of high-resolution protein structures. It is basically a log-weighted combination of the clashscore, percentage Ramachandran not favored and percentage bad side-chain rotamers, giving one number that reflects the crystallographic resolution at which those values would be expected. The MolProbity score is calculated as:

$$\begin{aligned}\text{MolProbity-score} = \; & 0.42574 * \ln(1+\mathit{clashscore}) \\ & + 0.32996 * \ln(1+\max(0, \mathit{rota}_{out} - 1)) \\ & + 0.24979 * \ln(1+\max(0, 100 - \mathit{rama}_{iffy} - 2)) + 0.5\end{aligned} \tag{8}$$

where *clashscore* is the number of unfavorable steric overlaps ≥ 0.4 Å, including Hydrogen atoms, and $\mathit{rota}_{out}$ and $\mathit{rama}_{iffy}$ are the percentages of the outliers of the side-

chain rotamers and the backbone torsion angles, respectively. Thus MolProbity is sensitive to steric clashes, rotamer outliers, and Ramachandran outliers. The weighting factors were computed from a log-linear fit to crystallographic resolution on a filtered set of PDB structures, so that a protein's MolProbity score is the resolution at which it's MolProbity score would be the expected value. Thus, lower MolProbity scores indicate more physically realistic models. Unlike the other scoring measures we use, MolProbity is not native-dependent that is; the native structure is not required to calculate it. This difference makes MolProbity score significantly distinct from the other scoring function used in this work.

## 2.4 Results And Discussion

We first present the analysis on the relative importance of the various energy terms used in 3Drefine approach. Then the overall results obtained by using 3Drefine refinement protocol have been evaluated on recent critical assessment of techniques for protein structure prediction (CASP) [81] in the refinement category along with a comparative analysis of 3Drefine against all the groups participated in CASP8 and CASP9 refinement experiments [41, 62] together with a recently published refinement method called FG-MD [31]. We also examine the local qualities of the CASP8 and CASP9 refinement targets in detail before and after refinement and compared that with the qualities of the native structures. Finally we assess the performance of 3Drefine on 107 CASP9 targets using the initial models generated by our structure prediction method, MULTICOM-REFINE [82] during the CASP9 experiment.

### 2.4.1 Effects Of Various Energy Terms

To examine the detailed effects of various energy terms; we split 3Drefine into six different runs:

1) $E_{bondlength} + E_{bondangle} + E_{torsion:}$ Minimization using only the bonded terms of ENCAD potential [76]. These are the first three terms of Eq. (1).

2) Ebondlength + Ebondangle + Etorsion + Etether: Tethering energy term has been added to the bonded terms of ENCAD potential.

3) $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{tether} + E_{hydrogenbonds}$: Explicit hydrogen bonding potential has been added to the ENCAD bonded terms and Tethering energy term.

4) $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{hydrogenbonds} + E_{kbpairwise}$: ENCAD bonded terms together with the hydrogen bonding potential and knowledge-based potential of mean force [64]. This is basically the total energy defined in Eq. (1) without the Tethering energy term.

5) $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{tether} + E_{kbpairwise}$: ENCAD bonded terms together with the Tethering energy term and knowledge-based potential of mean force [64]. The hydrogen bonding potential is omitted here from the total energy described in Eq. (1).

6) $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{hydrogenbonds} + E_{tether} + E_{kbpairwise}$: This is the total energy used in the 3Drefine refinement as presented in Eq. (1).

**Table 2.1**. Effects of various energy terms of 3Drefine refinement on CASP8 and CASP9 Refinement Targets*

| Round | # | Energy Terms[a] | GDT-HA[b] | TM-score[c] | RMSD[d] |
|---|---|---|---|---|---|
| CASP8 | 1 | $E_{bondlength} + E_{bondangle} + E_{torsion}$ | 6.843 | 9.309 | 3.003 |
| | 2 | $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{tether}$ | 6.890 | 9.318 | 3.003 |
| | 3 | $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{tether} + E_{hydrogenbonds}$ | 6.709 | 9.306 | 2.978 |
| | 4 | $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{hydrogenbonds} + E_{kbpairwise}$ | 6.896 | 9.318 | 2.998 |
| | 5 | $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{tether} + E_{kbpairwise}$ | 6.941 | 9.326 | 2.999 |
| | 6 | $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{hydrogenbonds} + E_{tether} + E_{kbpairwise}$ | **6.979** | **9.362** | **2.953** |
| CASP9 | 1 | $E_{bondlength} + E_{bondangle} + E_{torsion}$ | 7.298 | 10.355 | 4.353 |
| | 2 | $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{tether}$ | 7.358 | 10.372 | 4.344 |
| | 3 | $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{hydrogenbonds} + E_{tether}$ | 7.347 | 10.376 | 4.339 |
| | 4 | $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{hydrogenbonds} + E_{kbpairwise}$ | 7.262 | 10.344 | 4.350 |
| | 5 | $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{tether} + E_{kbpairwise}$ | 7.386 | 10.382 | 4.344 |
| | 6 | $E_{bondlength} + E_{bondangle} + E_{torsion} + E_{hydrogenbonds} + E_{tether} + E_{kbpairwise}$ | **7.388** | **10.388** | **4.339** |

[a] Combination of various energy terms used in 3Drefine refinement.
[b] Cumulative GDT-HA score.
[c] Cumulative TM-score.
[d] Average RMSD in Å.

* Numbers in bold represent the best score in each category.

Table 2.1 summarizes the average results on CASP8 and CASP9 refinement targets. First, the combination of only the bonded terms of ENCAD potential [76] exhibits slight

degrade in the quality of global topology as measured by cumulative GDT-HA score, cumulative TM-score, average RMSD score ( i.e. cumulative GDT-HA from 6.898 to 6.843, cumulative TM-score from 9.316 to 9.309 and average RMSD from 3.004 Å to 3.003 Å for CASP8 refinement targets and cumulative GDT-HA from 7.319 to 7.298, cumulative TM-score from 10.368 to 10.355 and average RMSD from 4.344 Å to 4.353 Å for CASP9 refinement targets).

After adding the Tethering energy term to the bonded terms of ENCAD potential [76], an apparent improvement is observed over the ENCAD bonded terms only with cumulative GDT-HA score, cumulative TM-score, average RMSD score of 6.890, 9.318 and 3.003 Å respectively for the CASP8 targets and 7.358, 10.372, 4.344 Å respectively for the CASP9 targets. Although, the quality of the models on an average are not improved as compared to the starting models for CASP8 targets except for a slight improvement in the cumulative TM-score, Tethering energy term proves to be a beneficial addition to the 3Drefine total energy potential. For CASP9 targets, only the addition of Tethering energy term demonstrates improvement over the starting models in terms of cumulative GDT-HA score (from 7.319 to 7.358) and cumulative TM-score (from 10.368 to 10.372), while the average RMSD remains unaltered.

Addition of the hydrogen-bonding potential to the bonded terms of ENCAD potential and Tethering energy term seems not to affect the quality of the models compared to bonded terms of ENCAD potential and Tethering potential in terms of cumulative GDT-HA score and cumulative TM-score. However, the average RMSD score has been reduced when compared to the starting models (i.e. average RMSD from 3.004 Å to 2.978 Å for CASP8 refinement targets and from 4.344 Å to 4.339 Å for CASP9 refinement targets).

To further test the effects of adding Tethering energy term, we execute the minimization by omitting the Tethering potential from the 3Drefine total energy term described in Eq. (1). The cumulative GDT-HA score has been reduced for both CASP8 (from 6.898 to 6.896) and CASP9 targets (from 7.319 to 7.262). An increase in average RMSD can also be observed for CASP9 targets (from 4.344 Å to 4.350 Å). For CASP8 targets, although the average RMSD is less than the initial models (2.998 Å), this is worse than the combination of bonded terms of ENCAD potential with Tethering energy term and hydrogen bonding potential.

To justify the use of explicit hydrogen bonding potential, we run 3Drefine minimization without the hydrogen bonding energy term in Eq. (1). The results shows further increase in the RMSD scores (2.999 Å for CASP8 refinement targets and 4.344 Å for CASP9 refinement targets).

Finally, we perform the minimization using all the energy terms as presented in Eq. (1). This approach has achieved the highest cumulative GDT-HA score, cumulative TM-score and lowest RMSD score for both CASP8 and CASP9 refinement targets.

### 2.4.2 Assessment Of 3Drefine On CASP8 Refinement Experiment

We evaluate the performance of 3Drefine refinement protocol on all the 12 targets in CASP8 refinement experiment and compared it with all other groups participating in CASP8 refinement category along with FG-MD [31] which is a recent work and did not participate in CASP8. For the assessment of the results, we gather the performance of all

the participating groups in CASP8 in terms of cumulative GDT-HA score, cumulative TM-score, average RMSD score and average MolProbity from the previously published works [31, 41] on the assessment of CASP8 refinement experiment.

**Table 2.2**. Comparison of 3Drefine results (with FG-MD and Top Five Groups in CASP8 and CASP9 Refinement Experiments)

| Round | Group Name | # Targets[a] | GDT-HA[b] | TM-score[c] | RMSD[d] | MolProbity[e] |
|---|---|---|---|---|---|---|
| CASP8 | FG-MD[f] | 12 | 6.979 | 9.362 | 2.953 | 2.575 |
| | 3Drefine | 12 | 6.932 | 9.329 | 2.994 | 2.349 |
| | Null[g] | 12 | 6.898 | 9.316 | 3.004 | 2.706 |
| | LEE | 12 | 6.86 | 9.195 | 3.117 | 2.613 |
| | LevittGroup | 12 | 6.701 | 9.16 | 3.047 | 2.875 |
| | FAMSD | 12 | 6.562 | 8.746 | 3.056 | 2.796 |
| | SAM-T08-human | 12 | 6.523 | 9.084 | 3.105 | 2.762 |
| | YASARARefine | 12 | 6.407 | 9.155 | 3.359 | 1.071 |
| CASP9 | 3Drefine | 14 | 7.388 | 10.388 | 4.339 | 2.101 |
| | FG-MD[f] | 14 | 7.387 | 10.386 | 4.331 | 2.183 |
| | ZHANG | 14 | 7.365 | 10.396 | 4.338 | 3.042 |
| | SEOK | 14 | 7.359 | 10.399 | 4.259 | 3.436 |
| | Null[g] | 14 | 7.319 | 10.368 | 4.344 | 2.521 |
| | FAMSD | 14 | 7.284 | 10.348 | 4.309 | 2.55 |
| | FAMS-MULTI | 14 | 7.284 | 10.348 | 4.44 | 2.55 |
| | KNOWMIN | 14 | 7.194 | 10.182 | 4.74 | 2.179 |

[a] Number of CASP targets in the Refinement Experiment.
[b] Cumulative GDT-HA score based on the first submitted model.
[c] Cumulative TM-score based on the first submitted model.
[d] Average RMSD based on the first submitted model with respect to the native structure in Å.
[e] Average MolProbity score based on the first submitted model.
[f] Not a participating group in CASP8 and CASP9 Refinement Experiment.
[g] The initial models for the CASP refinement experiment.

The groups have been ordered based on the cumulative GDT-HA score of refined models for all the 12 targets. Upper part of Table 2.2 summarizes the overall result of 3Drefine with FG-MD [31] and top five groups participating in CASP8. A complete list of the CASP8 groups is listed in appendix Table A1.

The Null group basically represents the starting model provided by the CASP organizers for refinement. Groups that perform worse than Null group have on average degraded the model rather than improving it. The results demonstrate other than FG-MD; 3Drefine is the method that could consistently drive the initial model closer to the experimental structure in terms of cumulative GDT-HA, TM-scores and average RMSD.

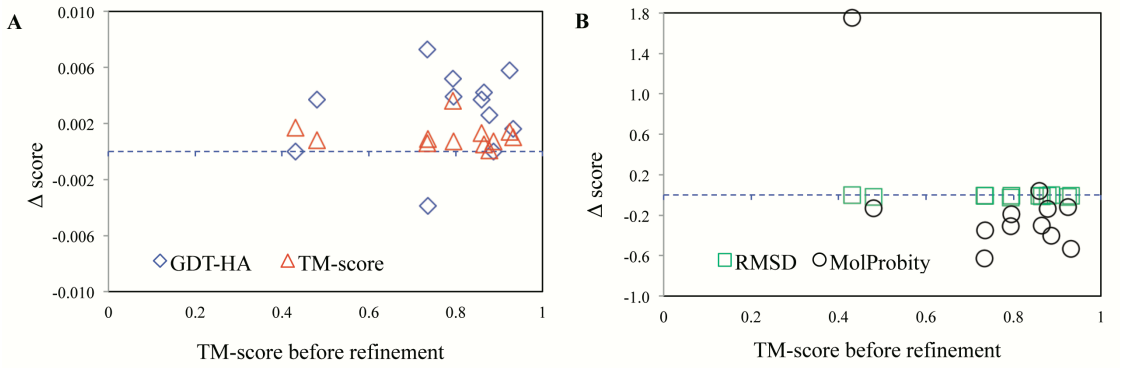


**Figure 2.1.** Changes in global and local structural qualities using 3Drefine on CASP8 Refinement Targets
(A) Scatter plot of changes in GDT-HA and TM-score. A positive change indicates the quality of the model of a target has been improved by refinement.
(B) Scatter plot of changes in RMSD and MolProbity-score. A negative change indicates the quality of the model of a target has been improved by refinement.

Overall, the cumulative GDT-HA and TM-score are 1.04% and 1.4% higher and RMSD is 0.123 Å lower than the second best LEE group participating in CASP8. The recent work, FG-MD, however, outperforms 3Drefine in these measures; although the performance is comparable. In terms of average MolProbity score, 3Drefine models

performed better than FG-MD, suggesting improved local qualities of the structures after refinement.

When compared with the starting models, 3Drefine exhibits consistent improvement in both global and local topologies of the initial structures. Out of the 12 CASP8 initial models, 3Drefine improves GDT-HA score for 9, TM-score for 12, RMSD for 11 and MolProbity score for 10 targets. In Figures 2.1A and 2.1B, we present the score changes (i.e. score after refinement – score before refinement) of the models refined by 3Drefine in terms of GDT-HA, TM-score, RMSD and MolProbity score against the TM-score of the starting model before refinement, which show that the qualities of the models refined by 3Drefine had been improved in most of the cases.

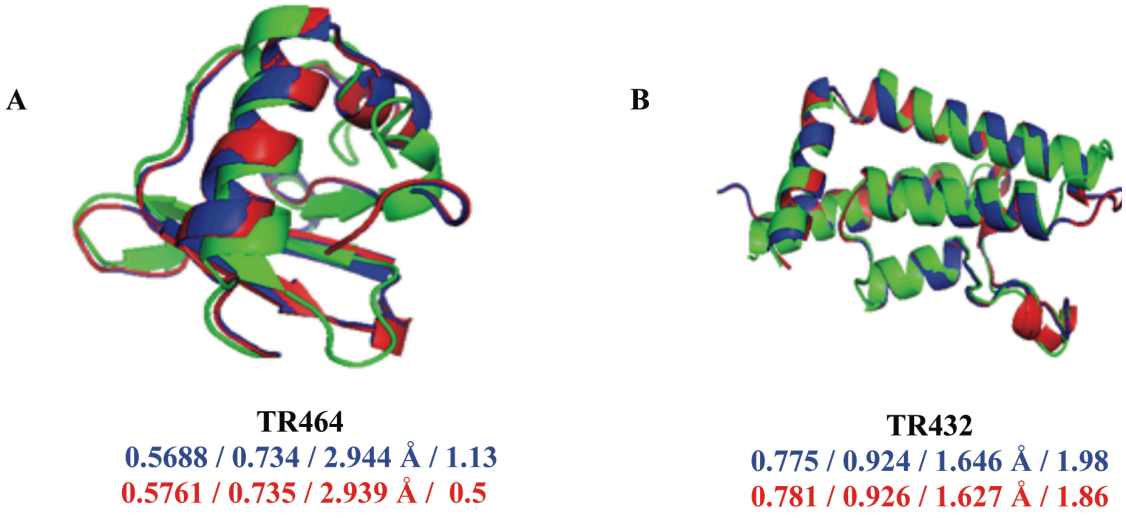


**Figure 2.2**. Structural superposition of initial model (blue) and refined model using 3Drefine (red) on the native structure (green) for two CASP8 Targets. The values under each model indicate GDT-HA, TM-score, RMSD and MolProbity score respectively before (blue) and after (red) refinement.
(A) Structural superposition for Target TR464.
(B) Structural superposition for Target TR432.

Two representative examples of CASP8 refinement have been presented in Figure 2.2. For the target TR464, 3Drefine refinement resulted in a 1.3% increase in GDT-HA score, a 0.1 % increase in TM-score and a 56 % decrease in MolProbity score. For the target TR432, the GDT-HA and TM-score improvement are 0.8 % and 0.4 % respectively while MolProbity improvement is 6 %.

### 2.4.3 Assessment Of 3Drefine On CASP9 Refinement Experiment

There were 14 targets available for refinement in CASP9 Refinement Experiment with length from 69 to 159 residues [62]. Along with the initial models and global distance test total score (GDT-TS), predictors were provided with hints about the focus regions, that is, groups of residues that need refinement. To ensure strict blind prediction, we do not use the hints or the starting GDT-TS score for 3Drefine run.

A summary of 3Drefine with top five CASP9 predictors ordered based on the cumulative GDT-HA score of the first model for all 14 targets have been listed in the lower part of Table 2.2. The results for the other groups have been adopted from the published works on CASP9 refinement assessment [31, 62] and a complete list of all the CASP9 groups has been presented in appendix Table A2.

The results show that 3Drefine, FG-MD, ZHANG and SEOK were able to consistently refine the staring model on the basis of GDT-HA, TM-score and RMSD score. However, the both ZHANG and SEOK models have MolProbity score higher than the initial model indicating degradation in the local qualities of the structures. The MolProbity

improvement for 3Drefine was 30.9% more than the best ZHANG group participating in CASP9. When compared to FG-MD, 3Drefine improves the MolProbity score by 3.8 %. Overall, 3Drefine protocol demonstrates consisted refinement of the initial model in terms of cumulative GDT-HA, cumulative TM-score, average RMSD and average MolProbity score; outperforming all the CASP9 predictors including the recent work FG-MD in terms of cumulative GDT-HA score.

In Figure 2.3A-3D we present the changes in GDT-HA, TM-score, RMSD and MolProbity score before and after refinement against the TM-score of the starting models for all the 14 CASP9 targets by 3Drefine and FG-MD. There are 9,13,11 and 12 cases when 3Drefine can improve the GDT-HA, TM-score, RMSD and MolProbity scores respectively; while FG-MD do so for 11, 9, 9 and 13 CASP9 targets.

Overall, the performance of 3Drefine is comparable to FG-MD in terms of its ability to enhance global qualities of the initial structure, that is, improvement on GDT-HA, TM-score and RMSD scores; although 3Drefine performs slightly better than FG-MD on CASP9 targets on these measures. However, with respect to the improvement of local qualities of structures, 3Drefine clearly outperforms FG-MD. The average MolProbity score of 3Drefine is 3.8 % better than that of FG-MD with an overall improvement of 16.6 % in MolProbity score over the starting models.

Figure 2.3E and 2.3F show two typical examples of refinement on CASP9 targets. For the target TR606, 3Drefine refinement resulted in a 3.1% increase in GDT-HA score, a 0.5 % increase in TM-score and a 21 % decrease in MolProbity score. There are 2.1 % and 0.6 % improvement of the GDT-HA and TM-score respectively and 8.1 % improvement in MolProbity score for the target TR622.

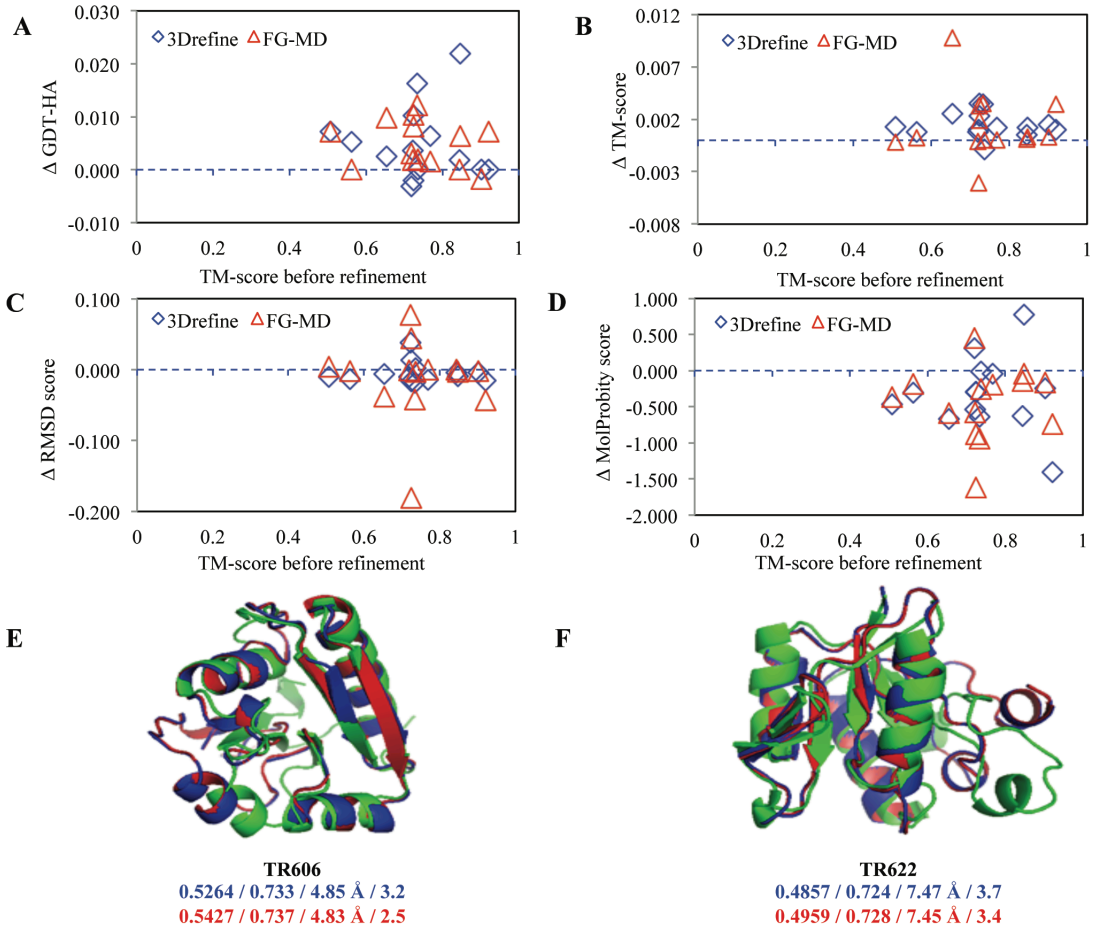


**Figure 2.3**. Changes in local and global qualities of CASP9 Refinement Targets
(A) Scatter plot for GDT-HA score changes for 3Drefine and FG-MD
(B) Scatter plot for TM-score changes for 3Drefine and FG-MD
(C) Scatter plot for RMSD-score changes for 3Drefine and FG-MD
(D) Scatter plot for MolProbity-score changes for 3Drefine and FG-MD
(E) Structural superposition of initial model (blue) and refined model using 3Drefine (red) on the native structure (green) for CASP9 target TR606.
(F) Structural superposition of initial model (blue) and refined model using 3Drefine (red) on the native structure (green) for CASP9 target TR624.
The values under each model indicate GDT-HA, TM-score, RMSD and MolProbity score respectively before (blue) and after (red) refinement.

## 2.4.4 A Closer Look At The Local Qualities Of CASP8 And CASP9 Refinement Targets

**Table 2.3**. MolProbity scores for CASP8 and CASP9 Refinement Targets

| Round | Target Name | Initial MolProbity[a] | Refined MolProbity[b] | Native MolProbity[c] |
|---|---|---|---|---|
| CASP8 | TR389 | 2.68 | 2.72 | 1.05 |
| | TR429 | 2.61 | 2.47 | 1.81 |
| | TR432 | 2.01 | 3.76 | 1.1 |
| | TR435 | 2.42 | 2.07 | 2 |
| | TR453 | 1.13 | 0.5 | 1.48 |
| | TR454 | 3.14 | 2.83 | 0.86 |
| | TR461 | 2.41 | 1.88 | 1.62 |
| | TR462 | 2.06 | 1.87 | 2.7 |
| | TR464 | 3.15 | 2.75 | 2.55 |
| | TR469 | 2.53 | 2.23 | 1.86 |
| | TR476 | 1.98 | 1.86 | 2.66 |
| | TR488 | 3.38 | 3.25 | 0.5 |
| **Average** | | **2.46** | **2.34** | **1.68** |
| CASP9 | TR517 | 1.4 | 1.36 | 1.02 |
| | TR530 | 0.9 | 1.67 | 0.64 |
| | TR557 | 1.5 | 1.49 | 1.14 |
| | TR567 | 1.4 | 0.77 | 3.6 |
| | TR568 | 1.5 | 1.19 | 0.56 |
| | TR569 | 0.7 | 1.01 | 2.05 |
| | TR574 | 3.6 | 2.93 | 0.5 |
| | TR576 | 3.7 | 3.16 | 1.46 |
| | TR592 | 3.5 | 2.09 | 1.94 |
| | TR594 | 2.9 | 2.65 | 1.01 |
| | TR606 | 3.2 | 2.56 | 0.81 |
| | TR614 | 4 | 3.7 | 1.67 |
| | TR622 | 3.7 | 3.4 | 1.42 |
| | TR624 | 1.9 | 1.43 | 2.33 |
| **Average** | | **2.42** | **2.1** | **1.44** |

[a] MolProbity score for the starting models.
[b] MolProbity score after 3Drefine refinement.
[c] MolProbity score for Native Structures.

Although the global structural measures like GDH-HA score, TM-score or the $C_\alpha$ RMSD scores provide the accuracy predicted protein models, they are primarily focusing on the

correctness of the backbone conformation of proteins and often fail to delve into finer atomic details of the predicted models. For instance, staggered $\chi$ angles are crucial details in estimating the qualities of protein structure [90]; but they are not considered in the global quality measures. Also the unfavorable steric clashes are strongly correlated with quality of a protein structure, with clashes reduced nearly to zero in the well-ordered parts of very high-resolution crystal structures [91]. In order to investigate these minute but vital aspects of models, we decide to perform a detailed analysis on the local structural qualities of the CASP8 and CASP9 refinement targets using MolProbity score - a single composite metric for local model quality.

All-atom contact, rotamers, and Ramachandran analysis are fundamental to the MolProbity structure-validation approach [92], which is widely accepted standard in macromolecular crystallography. CASP8 marks the first use of the MolProbity score for evaluation of non-experimental protein models. It is a very sensitive and demanding measure; attracting lot of attentions in serious works to assess the protein model qualities beyond $C_\alpha$ accuracy metrics [87].

Table 2.3 summarizes the MolProbity score for all CASP8 and CASP9 refinement targets. For each targets the score for initial model, score after 3Drefine refinement has been presented along with the MolProbity score for the native structures. It can be clearly seen that apart from a few targets (two targets in CASP8 and two targets in CASP9); the MolProbity score is always lower in the native structures when compared with the staring models. On an average, the MolProbity score for the initial structures in CASP8 and CASP9 are 2.46 and 2.42 respectively while the native structures have an average score of 1.68 and 1.44 respectively. The difference in the MolProbity score for the initial

models and the native structures undoubtedly demonstrate the need for the refinement of the local qualities of the starting structures in any refinement protocol.

Promisingly, 3Drefine exhibits improvement in the local qualities of the predicted protein models as measured by MolProbity score in CASP8 and CASP9 refinement targets. Apart from two targets in CASP8 (TR389 and TR432) and two targets in CASP9 (TR530 and TR569), the MolProbity score is always reduced compared to the starting models. Overall, the average MolProbity score for the refined models are 2.34 and 2.10 for CASP8 and CASP9 refinement targets respectively. Although the refined models are still far from achieving the average MolProbity score as the native structures, 3Drefine has certainly enhanced the local model qualities with respect to the initial structures.

### 2.4.5 Performance Of 3Drefine On 107 CASP9 Targets

To further assess the performance of 3Drefine on a large set of target models, we tested the refinement protocol on 107 CASP9 targets generated by our tertiary structure predictor MULTICOM-REFINE[82] that participated in the CASP9 experiment. We selected the first predicted model generated by MULTICOM-REFINE as the starting model for 3Drefine run for each of the 107 CASP9 targets. Similar to our testing strategy of 3Drefine for CASP8 and CASP9 refinement experiments, we performed the refinement in a blind mode, that is, without the knowledge of the native structure.

We observe a consistent improvement in the global qualities of the starting models after the refinement as measured by the GDT-HA, TM-score, and RMSD score. There were

59, 89 and 87 cases when 3Drefine brings the starting model closer to the native ones with respect to GDT-HA, TM-score and RMSD score respectively.

Overall, there was a 0.4 % increase in cumulative GDT-HA score and 0.1 % increase in cumulative TM-score for the refined models over the initial structures predicted by MULTICOM-REFINE for all the 107 CASP9 targets. The average RMSD of the refined models was 0.007 Å lower than the starting models.

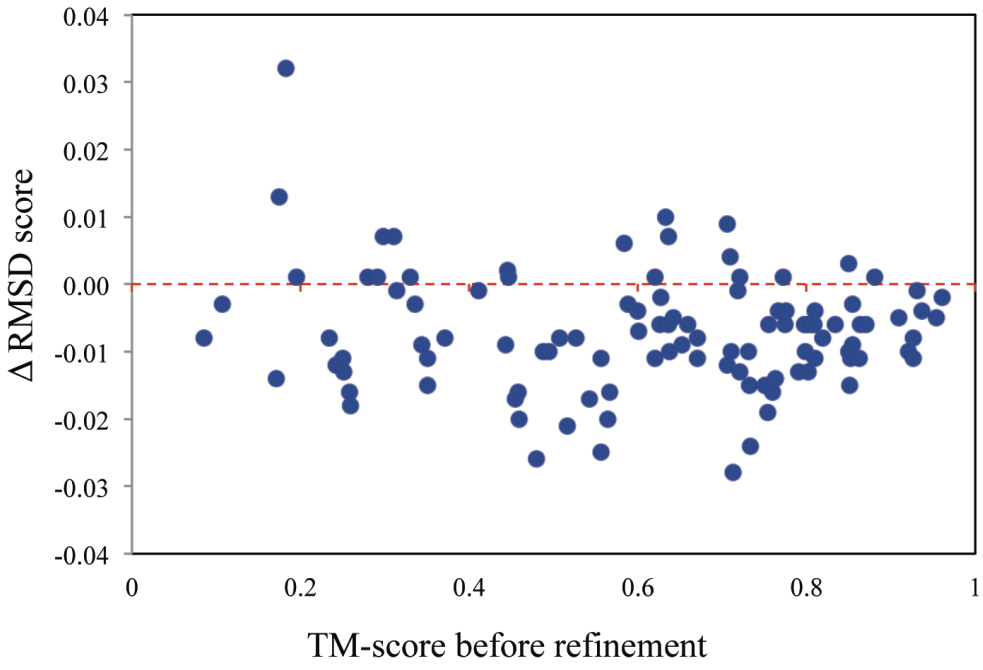


**Figure 2.4**. Scatter plot of RMSD changes for 107 CASP9 Targets (Initial models generated using MULTICOM-REFINE).

The changes of the RMSD score after refinement over the TM-score of the starting models has been shown as a scatter plot in Figure 2.4. In Figure 2.5A and 2.5B, we present the change of GDT-HA and TM-score before and after refinement against the initial TM-score for the 107 CASP9 targets. We controlled the initial TM-score Detailed

histograms of changes in the score have been shown in appendix Figure A1A and A1B. These results demonstrate the ability of 3Drefine protocol for consistent refinement of the CASP9 predicted models to bring it closer to its native structure in terms of global qualities of the structures. Most significant improvements have been observed when the TM-score of the starting model is > 0.5, that is, when the predicted models share the same fold with the native structure [89]. 3Drefine refinement results for the CASP8 and CASP9 refinement experiments along with the refinement of 107 CASP9 models are freely available at:

http://sysbio.rnet.missouri.edu/3Drefine/download.html.

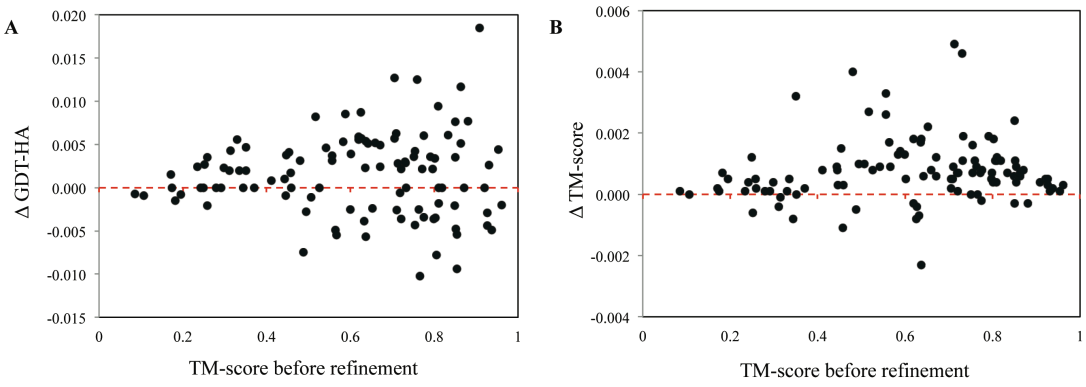


**Figure 2.5**. Refinement results for 107 CASP9 Targets using 3Drefine (Initial structures generated using MULTICOM-REFINE)
(A) GDT-HA score changes. A positive change indicates the quality of the model of a target has been improved by refinement
(B) TM score changes. A positive change indicates the quality of the model of a target has been improved by refinement.

## 2.5 Conclusion

Despite attracting constant attention by the researches, protein structure refinement problem remains a largely unsolved problem [93]. Because of the strong mutual association between the back-bone positioning and side-chain conformation of a protein model [94] simultaneous refinement of both the global topologies and local structural qualities of a protein structure is intended. Unfortunately, apart from a few promising works in the recent years [30, 31, 63, 64] majority of the structure refinement protocols fail to achieve this goal. Addressing this problem successfully would have major implications in resolving the bottleneck to apply computational protein structure prediction methods in structure-based drug design [95], protein docking [96] and prediction of biological functions based on protein structure [97]. As per the results of most recent critical assessment of techniques for protein structure prediction (CASP) refinement experiments, CASP8 and CASP9, there may be substantial room for improvement in the refinement category [41, 62].

In this work, we present a computationally efficient and reliable protocol for protein structure refinement, called 3Drefine. This method is a combination of two steps of minimization: Optimizing Hydrogen Bonding Network and Energy Minimization using a composite physics and knowledge based force fields, which is implemented within the MESHI [65] molecular modeling framework. It takes only few minutes (usually less than 5 minutes) of CPU time to refine protein models of usual length using 3Drefine.

3Drefine was tested in blind mode for CASP8 and CASP9 refinement targets in a completely automated manner, without using the knowledge about the information

provided by CASP organizers to focus on certain parts of the proteins for refinement. We observe that 3Drefine has the ability to consistently bring the model closer to the native structure. The models refined by 3Drefine have shown improvement of the global topologies of the starting models as measured by GDT-HA, TM-score and $C_\alpha$ RMSD to native structures as well as the local structure qualities as measured by the MolProbity score. The overall results of 3Drefine were better than or comparable to the other state-of the-art methods participating in CASP refinement category.

We also tested the performance on a large benchmark of 107 CASP9 targets by using MULTICOM-REFINE [82] as a structure prediction method to generate the initial structures. 3Drefine demonstrates consistent improvement in qualities of the initial models.

Although promising, the improvement in qualities of the starting models after 3Drefine refinement is often modest. This is the case with almost all other existing state-of-the-art refinement protocols primarily due to the limited accuracy of physics based empirical force fields used predominantly in the refinement methods. Broader samplings around the initial conformation of the protein using these force fields impose the risk of degrading the model quality instead of improving it. As the result, the refinement algorithms often rely on a more conservative strategy to sample locally around the starting structures producing improvement only in general physicality of the models rather than substantially improving the backbone positioning. Also, with the progress in the structure prediction pipelines, the qualities of the starting models are getting improved. Therefore, adopting more adventurous global search techniques at the cost of consistency that can improve the overall fold in the starting models are becoming less common amongst the

refinement pipelines. Even with the unadventurous strategies, the existing refinement protocols are often inconsistent as indicated by the results of CASP8 and CASP9 refinement experiment with majority of groups degrading the model qualities on an average. The unique nature of 3Drefine protocol is consistency. Around 80% of the times 3Drefine has improved the global qualities in the starting structures in CASP8 and CASP9 refinement targets. Also, the ability of 3Drefine to simultaneously improve the backbone positioning and local model qualities is encouraging. Future directions would be to use 3Drefine method in conjunction with some global search technique that can substantially improve the overall fold in the starting models together with the improvement in general physicality and local qualities of the models.

We conclude that 3Drefine can become a reliable and efficient method in protein structure refinement. The success of the protocol in improving accuracy of the initial models in a computationally inexpensive way for CASP refinement targets, where the initial model has been generated by a variety of structure prediction techniques, suggests that 3Drefine can be adopted as a final step in computational structure prediction pipeline.

# Chapter 3

# i3Drefine Software For Protein 3D Structure Refinement And Its Assessment In CASP10

## 3.1 Abstract

Protein structure refinement refers to the process of improving the qualities of protein structures during structure modeling processes to bring them closer to their native states. Structure refinement has been drawing increasing attention in the community-wide Critical Assessment of techniques for Protein Structure prediction (CASP) experiments since its addition in 8th CASP experiment. During the 9th and recently concluded 10th CASP experiments, a consistent growth in number of refinement targets and participating groups has been witnessed. Yet, protein structure refinement still remains a largely unsolved problem with majority of participating groups in CASP refinement category failed to consistently improve the quality of structures issued for refinement. In order to alleviate this need, we developed a completely automated and computationally efficient

protein 3D structure refinement method, i3Drefine, based on an iterative and highly convergent energy minimization algorithm with a powerful all-atom composite physics and knowledge-based force fields and hydrogen bonding (HB) network optimization technique. In the recent community-wide blind experiment, CASP10, i3Drefine (as 'MULTICOM-CONSTRUCT') was ranked as the best method in the server section as per the official assessment of CASP10 experiment. Here we provide the community with free access to i3Drefine software and systematically analyse the performance of i3Drefine in strict blind mode on the refinement targets issued in CASP10 refinement category and compare with other state-of-the-art refinement methods participating in CASP10. Our analysis demonstrates that i3Drefine is only fully-automated server participating in CASP10 exhibiting consistent improvement over the initial structures in both global and local structural quality metrics. Executable version of i3Drefine is freely available at http://protein.rnet.missouri.edu/i3drefine/.

## 3.2 Introduction

The biennial community-wide Critical Assessment of protein Structure Prediction (CASP) experiment aims to evaluate the progress and challenges in the state-of-the-art of protein structure modeling techniques, one of the fundamental problems in computational biology- prediction of the tertiary structure of protein from its sequence information. During the recent CASP experiments, encouraging and consistent progress have witnessed in template-based modeling (TBM) [33-36] or ab-initio (free-modeling; FM)

[37-40] folding of protein structures. The refinement category has been a recent addition to the CASP framework since CASP8, which aims to evaluate whether further improvement is possible to the best predictions made by contemporary structure prediction techniques. In the blind refinement experiment, predictors are given a starting structure evaluated by the organizers as the best submitted model during the structure prediction phase (TS category) along with the sequence information. Occasionally, some hints are also provided to aid the refinement like the focus regions during refinement or the accuracy of the starting structure.

Since its inclusion during CASP8, refinement category has been drawing increasing attention by the community. During recently concluded CASP10 refinement experiment, a 92% increase in the number of refinement targets and 39% increase in the number of participating groups have been observed compared to CASP9. This is not unexpected because a consistent and efficient refinement protocol can serve as a natural end step in almost all the contemporary structure prediction pipelines adding value to the already predicted structures through simultaneous improvement in backbone geometry and correction of local errors like irregular hydrogen bonding, steric clashes, unphysical bond length, unrealistic bond angles, torsion angles and side-chain $\chi$ angles. However, structure refinement has proven to extremely challenging as revealed in the assessment of refinement experiments during CASP8 and CASP9 [41, 42] with only a few participating groups were able to improve the model quality consistently. It should be noted, however, that CASP refinement category differs in a slight but significant way from refinement in the context of TBM [23, 27, 43-49] where the objective is to refine the best identified

template structure(s) to produce better quality prediction. In CASP, on the other hand, the starting models issued for refinement have already been refined by other structure prediction pipelines and judged to be the best among all the submitted models. Thus, attempts to improve qualities of these models would naturally impose more challenges and often the risk of degrading the model quality instead of improving it.

In view of the major difficulties in the field, we developed a consistent and computationally efficient refinement algorithm, called 3Drefine [98] by optimizing the hydrogen bonding network and atomic level energy minimization using a composite physics and knowledge-based force field. We participated in CASP10 refinement category with an iterative version of 3Drefine protocol, i3Drefine. As per the official CASP10 results released during CASP10 meeting in the form of assessors' presentation (http://predictioncenter.org/casp10/docs.cgi?view=presentations), i3Drefine was ranked as the single best refinement server method capable of consistent improvement in qualities of starting structures. The contribution of this article are two-fold: (1) Providing the community with access to a fast, accurate and freely downloadable executable version of refinement software which could be used to improve the qualities of the models coming from variety of protein structure prediction methods, or to act as the end-game strategy in a TBM pipeline and (2) evaluation of its performance in CASP10 refinement experiment to analyze the effectiveness of this method in a strict blind mode. Although CASP10 refinement category includes both human and server predictors, since i3Drefine is a fully automated server, this article will be mainly focused on the assessment of refinement in the context of automated server predictions.

## 3.3 Materials and Methods

### 3.3.1 i3Drefine Algorithm

i3Drefine is an iterative implementation of the energy minimization technique, 3Drefine for protein structure refinement. The details of 3Drefine protocol has been described in [98]. Here, we present a brief overview of 3Drefine algorithm.

3Drefine refinement protocol involves a two-step process: (1) Optimizing hydrogen bonding network and (2) atomic-level energy minimization using a combination of physics and knowledge based force fields; implemented using the molecular modeling package MESHI [65] . Given a starting structure for refinement, a combination of local geometry restraint and a conformational search is first performed in order to optimize the hydrogen bonding network. The optimized structure is called extended atomic model. Subsequently, 200,000 steps of energy minimization is employed on the extended atomic model using highly convergent limited memory Broyden–Fletcher–Goldfarb–Shannon (L-BFGS) [80] algorithm or until convergence to machine precision using a customized all-atom force field. The force field consists of a combination of physics based and knowledge based terms. The energetic contributions of the bonded interactions described in ENCAD potential [76] (bond length, bond angle, and torsion angle) along with tethering term of the $C_\alpha$ and $C_\beta$ atoms [98] constitute the physics-based part while atomic pairwise potential of mean force [24] and explicit hydrogen bonding potential [31] account for the knowledge-based terms. A detailed analysis of the relative importance of

these energy terms has been presented in the published work of 3Drefine [98]. The energy-minimized model is the refined model.

In i3Drefine, we use an iterative version of 3Drefine method. In order to escape from the local minima and move closer to the native structure, the starting model is minimized using 3Drefine protocol and the resulting refined model is again processed by the same method. This iteration is done five times to generate five refined models for the starting structure. Because 3Drefine invokes restrained backbone flexibility during energy minimization due to the inclusion of the knowledge-based terms in the all-atom force field, such an iterative scheme is effective. Furthermore, because of the computationally inexpensive nature of 3Drefine protocol, this iterative strategy does not provide significant computational overhead in i3Drefine pipeline consuming only a few minutes (typically less than 15 minutes) to generate five refined structures at a 2.4 GHz CPU.

### 3.3.2 Programming Language, Platform And External Programs

The core of i3Drefine is developed in Java (http://www.java.com/en/) on top of MESHI [65] software package and the command-line interface to perform the refinement is developed in Perl programming language (http://www.perl.org/). For a seamless installation and usage of i3Drefine, a Java version 6.0 or above and Perl version 5.8.8 or above is recommended. Also, since some of the energy terms in the customized force

fields require the secondary structure assignment of the starting structure for accurate calculations, DSSP program [77] needs to be used in conjunction with i3Drefine. The detailed installation instructions along with typical example of using i3Drefine have been provided in the user manual file supplied with the software. i3Drefine has been tested on 64-bit Linux based platform. However, because of the platform independent nature of Java and versatile platform support of Perl, it can be fairly easily modified to run for Windows or Mac OSX platforms.

### 3.3.3 Metrics Used For Evaluation

We evaluate the quality of the structural refinement using both global and local measures. We focus on GDT-TS [83] and RMSD [85] score to measure of the global positioning of $C_\alpha$ atoms. Global distance cutoff sidechain (GDC-SC) [34] has been used as a global quality metric for sidechain positioning. To assess the local qualities of the models, we use MolProbity score [86] as a local measure of physical correctness of a structure and SphereGrinder [42] as a local all-atom measure of structural similarity. Finally we use a recently introduced contact area difference (CAD) score [99] which quantifies the differences between physical contacts in the models before and after refinement with respect to their native structures.

#### 3.3.3.1 GDT-TS

GDT-TS [83] is a global quality measure of the correct positioning of backbone based on multiple superpositions of the predicted and experimental structure. It counts the average percentage of residues with $C_\alpha$ atom distance from the native structure residues below 1, 2, 4, and 8 Å, respectively, after optimal structure superposition. GDT-TS ranges from (0, 1) with higher value indicating better accuracy.

#### 3.3.3.2 RMSD

Similar to GDT-TS, RMSD [85] is a global measure of the correct positioning of the $C_\alpha$ atoms. However, RMSD is based on a single superposition lacking any kind of distance cutoffs. Hence, RMSD and GDT-TS is weekly correlated. Furthermore, unlike GDT-TS, a lower RMSD value indicates that the predicted structure is close to its native state.

#### 3.3.3.3 GDC-SC

GDC-SC [34] has been used as a global quality metric for sidechain positioning. Unlike GDT-TS, which is focused on $C_\alpha$ atoms, GDC-SC use a single characteristic atom near the end of each sidechain. Also, 10 different superpositions with different weighting schemes are employed to calculate GDC-SC.

#### 3.3.3.4 MolProbity

In order to evaluate the physical realism and the local errors, we use MolProbity [86] – a single and composite score to measure local model quality. The MolProbity score denotes

the expected resolution of the protein model with respect to standard experimental structures and therefore, lower MolProbity score indicates more physically realistic model.

### 3.3.3.5 SphereGrinder

We use SphereGrinder to measures the local environment around each residue which was used in the refinement assessment of CASP9 [42]. SphereGrinder is based on an all-atom RMSD fit between the experimental and predicted structures using a sphere constructed by considering the set of atoms within 6 Å of the $C_\alpha$ atoms for each residue in experimental structure.

### 3.3.3.6 CAD-AA

CAD score [99] is a newly introduced quality metric which is based on contact area difference between predicted and experimental structure, thereby directly reflecting interactions within the protein structure. The contact area is calculated based on a protein structure tessellation approach [100] and normalized between [0, 1] with higher value indicating better structure. We use the all-atom version of the CAD score, namely, CAD-AA.

### 3.3.3.7 Normalizing And Overall Quality Score

Higher value of GDT-TS, GDC-SC, SphereGrinder and CAD-AA scores indicate better models while lower values RMSD and MolProbity scores represent better models. In order to effectively compare the degree of refinement between different groups or targets,

a single overall quality score is essential. We use a robust version of Z-score based on median absolute difference (MAD) of the changes in quality of the models induced through refinement. This is a slightly modified approach used in refinement assessment during CASP9 [42].

The difference in the model quality is first calculated to get the delta quality score for a given quality metric (e.g. GDT-TS).

$$\delta_Q(r) = Q(r) - Q(s) \quad (1)$$

where Q(r) and Q(s) denote the quality score for refined and starting structures respectively corresponding to quality measure Q.

For a given target, we calculate the MAD using:

$$MAD_{\delta_Q} = median(|\delta_Q(r) - median(\delta_Q)|) \quad (2)$$

where median($\delta_Q$) denotes the median of the delta score for the corresponding quality metric and |.| is the absolute value. The robust Z-score is then calculated as:

$$Z_{r,Q} = \frac{\delta_Q(r) - median(\delta_Q)}{1.486 \times MAD_{\delta_Q}} \quad (3)$$

The factor 1.486 scales the MAD to be same as standard deviation of a normal distribution.

Finally, a weighted average of Z-score is taken for all different quality metrics to combine the results of all six scores into a single score, called Q-score.

$$Q_{overall,r} = \frac{5 \times Z_{r,GDT-TS} - Z_{r,RMSD} + Z_{r,GDC-SC} + Z_{r,MP} + Z_{r,SpGr} + Z_{r,CAD-AA}}{10} \quad (4)$$

In this scoring scheme, GDT-TS is given a weight of 5, which makes half of the overall score and other five metrics makes the other half. Although this procedure is arbitrary, it emphasizes the improvement in backbone positioning as judged by GDT-TS score, a widely used metric by CASP assessors, compared to other measures.

## 3.4 Results and Discussion

The fully automated i3Drefine software was first blindly tested in CASP10 refinement experiment, 2012 with the group name MULTICOM-CONSTRUCT (Server group 222). Since then, we systematically evaluate its performance using global and local quality metrics like GDT-TS, RMSD, GDC-SC, MolProbity, SphereGrinder and CAD-score and perform comparative analysis of i3Drefine against all the groups participating in CASP10 refinement category. Here, we first summarize the targets offered for refinement during CASP10 refinement experiment along with the measures of the initial quality. Secondly, we present the automated server groups participating in CASP10 refinement category and introduce a pseudo group called "Void" as a control. Thirdly, we assess the overall degree of refinement produced by i3Drefine in a strict blind mode. Fourthly, a comparison of i3Drefine against the state-of-the-art refinement server methods participating in CASP10 has been presented along with head-to-head comparison of the scores and their statistical significance. During CASP10, each predictor was asked to

submit up to five predictions while ranking submissions from best to worst. We, therefore, perform one set of analysis using the first submitted model, which is the best prediction as per the ranking from the predictor. However, because predictors often fail to correctly rank their submissions, we present a second set of analysis by selecting the best prediction (as evaluated by our overall quality score) from each group for each target. The comparison between the first and the best predicted models by i3Drefine also reveals the advantages of the iterative version of our refinement method (i3Drefine) over the non-iterative version (3Drefine). Finally, we compare i3Drefine with the top five non-server (human) methods and discuss the added benefits of human predictors and the possibility of adopting them in computational structure prediction pipelines.

### 3.4.1 Targets Used For Refinement In CASP10

Table 3.1 summarizes the targets issued for refinement in CASP10 and the measures of the initial quality of these targets. The occasional "hints" provided by the organizers to focus on certain segment(s) of the structures during refinement has also been reported. These are the starting models for refinement and were chosen from the top submissions during the structure prediction category. These models, therefore, represent one of the best predicted structures submitted by the community for each target and intuitively, consistent refinement of these structures is a nontrivial task.

**Table 3.1**. Summary of CASP10 refinement targets

| # | Target | Residues | Method | GDT-TS | RMSD (Å) | GDC-SC | MP[a] | SG[b] | CAD-AA | Focus[c] |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | TR644 | 141 | X-ray | 0.8422 | 2.712 | 0.4346 | 2.49 | 0.7518 | 0.69 | - |
| 2 | TR655 | 175 | NMR | 0.6871 | 4.654 | 0.2853 | 3.83 | 0.5143 | 0.58 | 6-20; 51-64 |
| 3 | TR661 | 185 | X-ray | 0.800 | 2.743 | 0.375 | 1.11 | 0.7135 | 0.67 | - |
| 4 | TR662 | 75 | NMR | 0.8267 | 2.031 | 0.3364 | 2.42 | 0.7600 | 0.67 | - |
| 5 | TR663 | 152 | X-ray | 0.6908 | 3.372 | 0.2626 | 4.05 | 0.7697 | 0.67 | 53-78; 141-181 |
| 6 | TR671 | 88 | X-ray | 0.5568 | 7.716 | 0.1158 | 3.68 | 0.4432 | 0.59 | - |
| 7 | TR674 | 132 | X-ray | 0.8523 | 3.444 | 0.4417 | 2.99 | 0.7424 | 0.69 | - |
| 8 | TR679 | 199 | X-ray | 0.7186 | 3.949 | 0.3076 | 1.15 | 0.5226 | 0.6 | 25-45; 146-156; 187-197 |
| 9 | TR681 | 191 | X-ray | 0.7827 | 2.273 | 0.3274 | 2.89 | 0.6387 | 0.64 | - |
| 10 | TR688 | 185 | X-ray | 0.7838 | 2.524 | 0.4249 | 1.77 | 0.7730 | 0.67 | - |
| 11 | TR689 | 214 | X-ray | 0.8773 | 1.660 | 0.4202 | 3.18 | 0.8738 | 0.72 | - |
| 12 | TR696 | 100 | X-ray | 0.7075 | 3.519 | 0.2631 | 2.97 | 0.5000 | 0.58 | - |
| 13 | TR698 | 119 | X-ray | 0.6471 | 4.653 | 0.2568 | 2.73 | 0.6555 | 0.63 | 17-35; 90-100 |
| 14 | TR699 | 225 | X-ray | 0.8411 | 2.211 | 0.3361 | 2.77 | 0.7733 | 0.66 | - |
| 15 | TR704 | 235 | X-ray | 0.6989 | 2.78 | 0.2325 | 2.89 | 0.7319 | 0.64 | - |
| 16 | TR705 | 96 | X-ray | 0.6458 | 4.709 | 0.2211 | 3.63 | 0.3750 | 0.52 | - |
| 17 | TR708 | 196 | X-ray | 0.8648 | 4.630 | 0.4551 | 2.65 | 0.8214 | 0.71 | - |
| 18 | TR710 | 194 | X-ray | 0.7487 | 2.440 | 0.3628 | 0.50 | 0.7732 | 0.72 | - |
| 19 | TR712 | 186 | X-ray | 0.9261 | 1.992 | 0.5515 | 2.69 | 0.8817 | 0.77 | 80-89; 116-129; 141-155 |
| 20 | TR720 | 198 | X-ray | 0.5783 | 8.515 | 0.2558 | 1.33 | 0.4697 | 0.58 | - |
| 21 | TR722 | 127 | X-ray | 0.5709 | 4.422 | 0.1614 | 0.88 | 0.8976 | 0.72 | - |
| 22 | TR723 | 131 | X-ray | 0.8511 | 2.232 | 0.3772 | 2.21 | 0.8473 | 0.68 | - |
| 23 | TR738 | 249 | X-ray | 0.9006 | 1.396 | 0.5036 | 2.38 | 0.9398 | 0.75 | 17-35; 90-100 |
| 24 | TR747 | 90 | X-ray | 0.825 | 1.956 | 0.3796 | 1.95 | 0.6778 | 0.63 | - |
| 25 | TR750 | 182 | X-ray | 0.7679 | 2.125 | 0.348 | 2.49 | 0.7967 | 0.67 | - |
| 26 | TR752 | 148 | X-ray | 0.9037 | 1.495 | 0.4305 | 1.52 | 0.7973 | 0.71 | 41-50; 100-110; 125-128 |
| 27 | TR754 | 68 | NMR | 0.7794 | 2.410 | 0.1997 | 2.56 | 0.8235 | 0.65 | - |

[a] MolProbity scores of the starting structures.
[b] SphereGrinder scores of the starting structures.
[c] The numbers indicate the range of focus residues as suggested by CASP10 organizers.

## 3.4.2 Server Groups Participating In CASP10 Refinement Category

A total of fifty groups participated in CASP10 refinement experiment including both human and server predictors. Thirteen groups took part as fully automated server predictors. The server predictors were given a three days deadline to submit the refined

structures to the prediction center as opposed to a three weeks deadline offered for the human predictors.

**Table 3.2.** List of server groups participating in CASP10 refinement category

| # | Group # | Group Name[a] | Targets attempted | Total submitted models | First submitted model[b] |
|---|---|---|---|---|---|
| 1 | 006 | **MUFOLD-QA** | 22 | 110 | 22 |
| 2 | 028 | **YASARA** | 18 | 18 | 18 |
| 3 | 103 | PconsM | 2 | 10 | 2 |
| 4 | 108 | **PMS** | 27 | 135 | 27 |
| 5 | 124 | PconsD | 2 | 10 | 2 |
| 6 | 175 | **FRESS_server** | 27 | 135 | 27 |
| 7 | 179 | Lenserver | 2 | 10 | 2 |
| 8 | 198 | **chuo-fams-server** | 27 | 27 | 27 |
| 9 | 222 | **MULTICOM-CONSTRUCT**[c] | 27 | 135 | 27 |
| 10 | 238 | **chuo-repack-server** | 26 | 26 | 26 |
| 11 | 286 | Mufold-MD | 1 | 5 | 1 |
| 12 | 292 | Pcons-net | 2 | 10 | 2 |
| 13 | 424 | **MULTICOM-NOVEL** | 27 | 135 | 27 |

[a] Group name in bold indicates the group has attempted more than 50% of refinement targets.
[b] Models submitted with a Model ID of one.
[c] CASP10 group name for i3Drefine.

In Table 3.2, we summarize the server groups participating in CASP10 along with the number of predictions submitted by each predictor. The performance of fully automated i3Drefine method (group name MULTICOM-CONSTRUCT) can be directly compared to these methods on the CASP10 refinement targets. This would enable us to assess the ability of i3Drefine protocol with state-of-the-art automated refinement methods in a strict blind mode. Groups attempting more that 50% of the targets have been highlighted in bold in Table 3.2.

As a control, we created a pseudo group called “Void” group. This group represents the starting model provided by the CASP organizers for refinement. We judge the success and degree of refinement with respect to the ‘Void’ group. Groups that perform worse

than Void group have on average degraded the quality of starting structures rather than improving it.

### 3.4.3 Overall Performance Of i3Drefine In CASP10 Blind Refinement Experiment

Figure 3.1 shows the distribution of change in model quality relative to the starting model as judged by the score difference in six quality metrics for all submitted model by i3Drefine method for all CASP10 refinement targets. Positive changes in GDT-TS, GDC-SC, SphereGrinder and CAD-AA scores represent refinement success whereas negative changes in RMSD and MolProbity scores indicate a failure in refinement. In Figure 3.1, the regions shaded in black indicate improvement in the corresponding quality measure with the numbers above these regions representing the percentage of refinement successes while the regions without shading indicate degradation in the model quality and the numbers specify the percentage of failures in refinement. While for most metrics, the number of improvements significantly outnumbered number of failures, the improvement is typically modest in nature. For example, refinement successes outnumber failures by more than a factor of three in global position of the backbone atoms as judged by GDT-TS and RMSD scores and global quality of sidechain positioning as measured by GDC-SC score. While most of ΔGDT-TS, ΔRMSD and ΔGDC-SC scores lie within ~ ±4%, the distributions are skewed towards improvement. Highly consistent improvement has also been observed in the local quality measures like ΔSphereGrinder and ΔCAD-AA scores

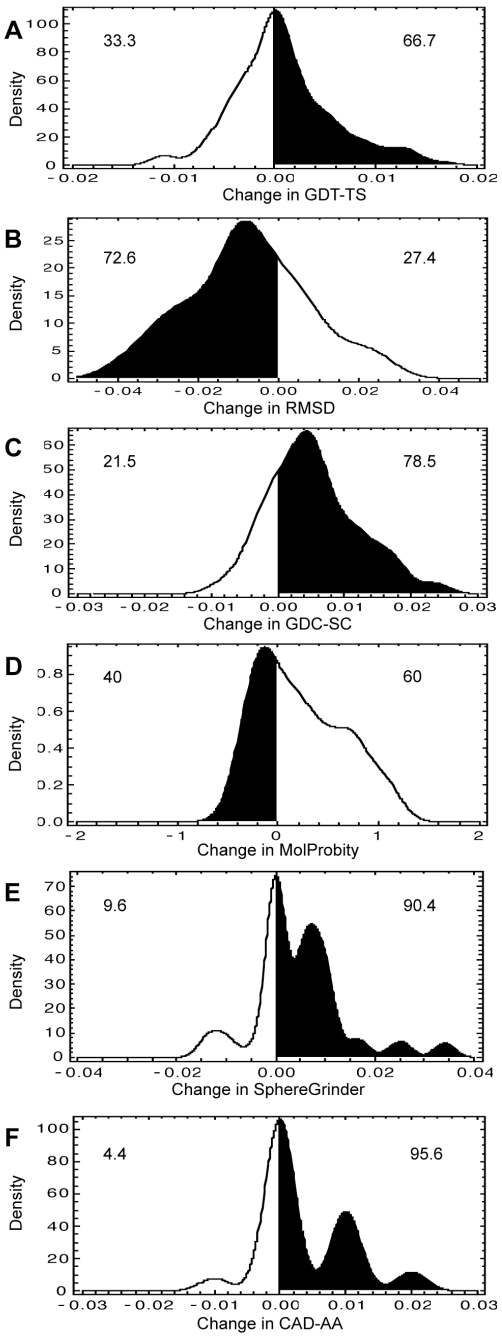


**Figure 3.1**. Distribution of i3Drefine refinement for all submitted structures.

Distributions of change in quality scores after i3Drefine refinement are shown for these metrics: (A) GDT-TS, (B) RMSD, (C) GDC-SC, (D) MolProbity, (E) SphereGrinder and (F) CAD-AA. Regions shaded in black indicate improvement over the starting model. The numeric values are the percentage of times the structures were made better or worse for each metric.

and the distributions are highly skewed towards success with over 90% success. However, for MolProbity score, there are more failures than success and the distribution is marginally skewed towards failure. The distributions in Figure 3.1 are multimodal, which indicate that not all targets are equally easy to refine and the degree of refinement vary with the difficulty of targets.

In Figure 3.2, we examine the relationship between the starting score of any of the quality measures and the ability of i3Drefine to improve the starting model. Although, it is difficult to infer a conclusive correlation between them with only 27 targets, some interesting trends can be observed. For example, most of the starting structures have quite accurate backbone positioning with only 7 out of 27 targets have RMSD score more than 4Å and GDT-TS less than 0.7. For these moderate-accuracy targets, i3Drefine always improves the backbone quality by increasing GDT-TS score and reducing RMSD score. For the more accurate starting structures with RMSD ~ 2Å, the RMSD distribution is skewed towards improvement. However, there are approximately as many improvements as failures in GDT-TS score for high-accuracy targets (GDT-TS more than 0.8). The global quality of sidechains, as measured by GDC-SC varies from 0.1 to 0.6 indicating that the starting structure set comprises a wide variety in terms of accuracy of sidechain positioning, although most of the targets are in the range of 0.3 to 0.5. Promisingly, i3Drefine consistently improves the GDC-SC score irrespective of the quality of starting structures. When the initial model has less accurate local quality as measured by MolProbity (MolProbity score is more than 2), we observe consistent improvement in MolProbity. However, i3Drefine almost always increases MolProbity score indicating

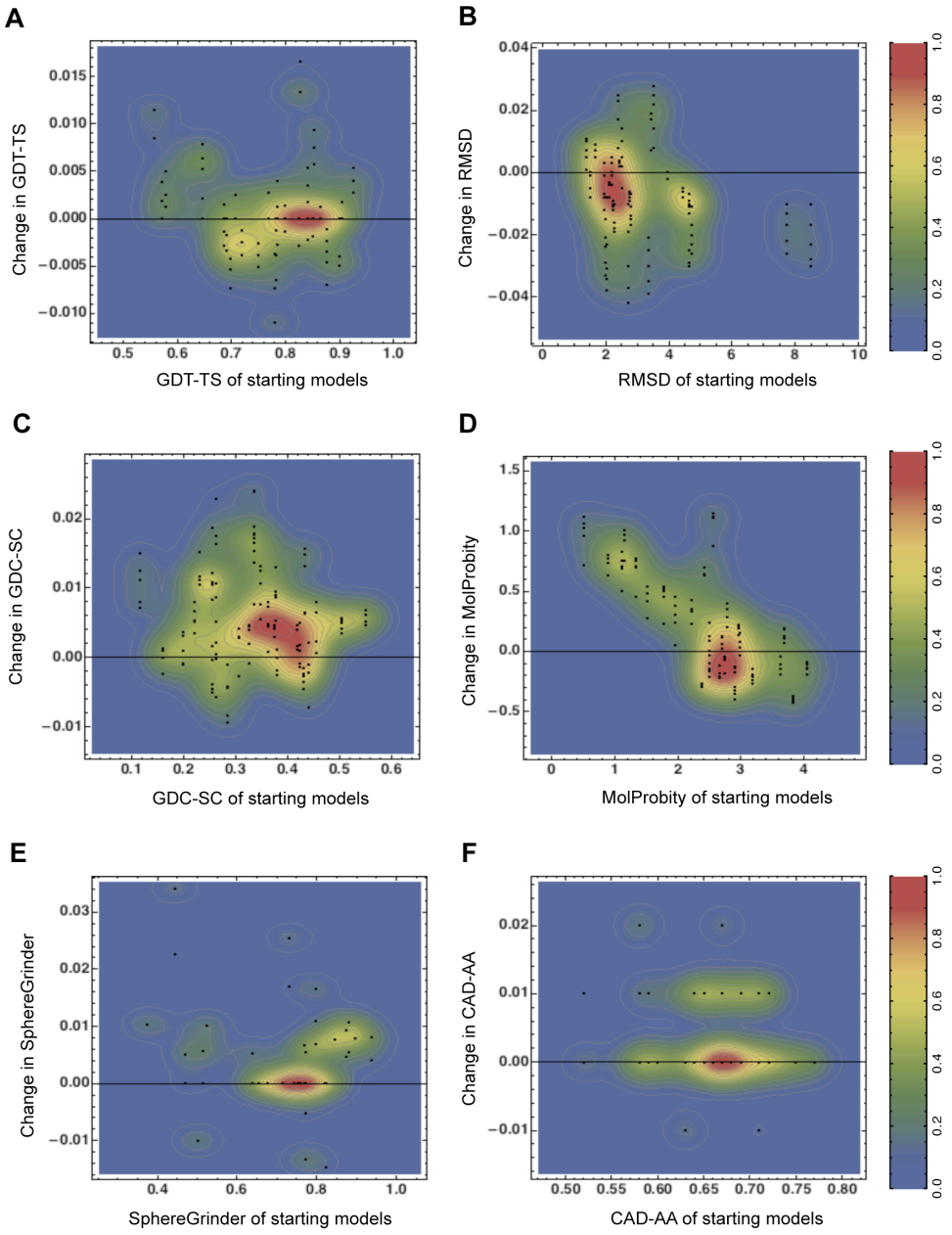


**Figure 3.2**. Distributions of score changes with respect to the quality of starting structures.

Relationships between changes in quality scores and the quality of the starting models are shown for these metrics: (A) GDT-TS, (B) RMSD, (C) GDC-SC, (D) MolProbity, (E) SphereGrinder and (F) CAD-AA. The Black points indicate the actual data points while the contours are filled with colours that vary from blue for low density to red for high density. The colour function has been scaled between 0 and 1 and the legends are shown on the right.

degradation in local model quality when MolProbity score is less than 2. For other local quality measures like SphereGrinder and CAD-AA, we observe a modest but consistent improvement in the model quality across all target difficulty. In short, more consistent and simultaneous improvements both in global and local quality measures have been observed for moderately accurate targets than high-accuracy targets.

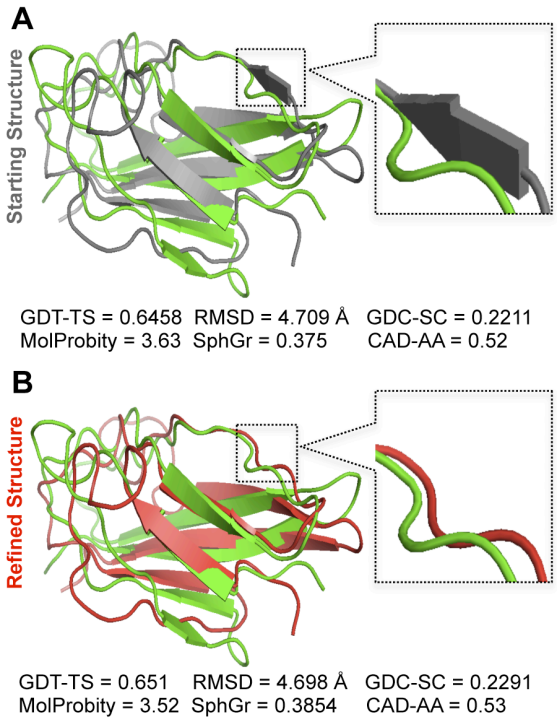


**Figure 3.3**. Example of i3Drefine refinement for CASP10 target TR705.
(A) Structural superposition of initial model (grey) on native structure (green).
(B) Structural superposition of refined model using i3Drefine (red) on native structure (green).
The values of the quality measures before and after refinement have been reported under the models. The black dotted square highlights the region with prominent structural improvements and a closer look of the change is shown in the right.

A representative example of refinement has been presented in Figure 3.3 for CASP10 refinement target TR705. i3Drefine refinement results in GST-TS, GDC-SC,

SphereGrinder and CAD-AA scores to increase from 0.6458, 0.2211, 0.375 and 0.52 to 0.651, 0.2291, 0.3854 and 0.53 respectively. The RMSD and MolProbity score decreases from 4.709 Å and 3.53 to 4.698 Å to 3.52 respectively. Clearly, a modest yet consistent improvement in all quality measures has been observed. More pronounced structural improvement in terms of backbone positioning has been observed around residue 58 where a disoriented strand region is rearranged to a coil, thereby bringing the refined model closer to the native state.

### 3.4.4 Comparison Of i3Drefine With Other Server Predictors Participating In CASP10

We compare the performance of i3Drefine with the thirteen server predictors participating in CASP10 refinement category based on the first submitted model and the best submitted model as judged by our overall quality score, $Q_{overall}$. It can be noticed form Table 3.2 that some of the predictors attempted very few targets and only eight groups (including i3Drefine) submitted prediction for more than 50% of targets (i.e. more than 13 targets). Although we have taken into account all the submitted models by every group while performing our analysis, we choose to focus on these eight predictors for a fair comparison between them. To compare predictors with a single score, we have computed the sum of $Q_{overall}$ for each predictor and ranked groups based on that.

Upper part of Table 3.3 summarizes cumulative change in all the quality measures with respect to the starting structures (represented as 'Void' group) for eight server predictors. The groups have been ordered based on the cumulative $Q_{overall}$ score for all the submitted

targets. The results demonstrate except MolProbity score, i3Drefine improves all the quality measures in terms of cumulative change with respect to the starting structures.

**Table 3.3.** Cumulative improvement relative to starting model for the top server groups in CASP10 refinement experiment*

| Selection | Group Name | ΔGDT-TS | ΔRMSD (Å) | ΔGDC-SC | ΔMP[a] | ΔSG[b] | ΔCAD-AA | Q_overall[c] |
|---|---|---|---|---|---|---|---|---|
| First Model | MULTICOM-CONSTRUCT[d] | 0.036 | -0.165 | 0.120 | 3.93 | 0.096 | 0.120 | 12.69 |
| | Void | 0.000 | 0.000 | 0.000 | 0.00 | 0.000 | 0.000 | 10.91 |
| | chuo-fams-server | 0.003 | 0.074 | -0.191 | 5.05 | -0.803 | -0.060 | 5.71 |
| | YASARA | -0.372 | -2.138 | 0.188 | -27.75 | -0.800 | 0.270 | 2.79 |
| | MULTICOM-NOVEL | -0.490 | 27.479 | -0.294 | 11.72 | -0.253 | -0.090 | -0.97 |
| | PMS | -0.534 | 9.606 | 0.097 | 9.73 | -0.624 | -0.010 | -5.97 |
| | chuo-repack-server | -0.364 | 1.939 | -0.523 | 13.86 | -0.439 | -0.430 | -8.27 |
| | FRESS_server | -1.574 | 10.155 | -1.102 | 32.79 | -1.512 | -0.930 | -39.53 |
| | MUFOLD-QA | -2.194 | 140.087 | -1.604 | 20.37 | -3.121 | -1.340 | -109.18 |
| Best Model | MULTICOM-CONSTRUCT[d] | 0.068 | -0.224 | 0.151 | 4.25 | 0.124 | 0.140 | 13.96 |
| | PMS | -0.162 | -2.091 | 0.382 | 8.29 | -0.025 | 0.170 | 13.72 |
| | Void | 0.000 | 0.000 | 0.000 | 0.00 | 0.000 | 0.000 | 10.91 |
| | MULTICOM-NOVEL | -0.220 | 14.010 | -0.105 | 9.79 | -0.059 | 0.010 | 7.49 |
| | chuo-fams-server | 0.003 | 0.074 | -0.191 | 5.05 | -0.803 | -0.060 | 5.71 |
| | YASARA | -0.372 | -2.138 | 0.188 | -27.75 | -0.800 | 0.270 | 2.79 |
| | chuo-repack-server | -0.364 | 1.939 | -0.523 | 13.86 | -0.439 | -0.430 | -8.27 |
| | FRESS_server | -0.926 | 4.787 | -0.648 | 30.10 | -0.865 | -0.520 | -21.75 |
| | MUFOLD-QA | -1.968 | 31.832 | -1.289 | 13.17 | -2.436 | -0.980 | -62.30 |

* The values for all quality metrics represent the cumulative change relative to the starting structures for all targets.
[a] Cumulative change in MolProbity score.
[b] Cumulative change in SphereGrinder score.
[c] Sum of overall quality score for all targets.
[d] CASP10 group name for i3Drefine.

In Figure 3.4, we present the distributions of changes in model quality relative to the starting models for the eight server predictors as measured by six quality metrics. Similar to Figure 3.1, the regions shaded in black in Figure 3.4 correspond to refinement successes while the regions without shading indicate failures in refinement. We also report the percentage of successes and failures for each quality measures. The distributions for each predictor are multimodal due to variations in the quality of the starting models. Also, the degree of change in the quality score varies between predictors and type of quality metric. We, therefore, choose to maximally cover the range of score

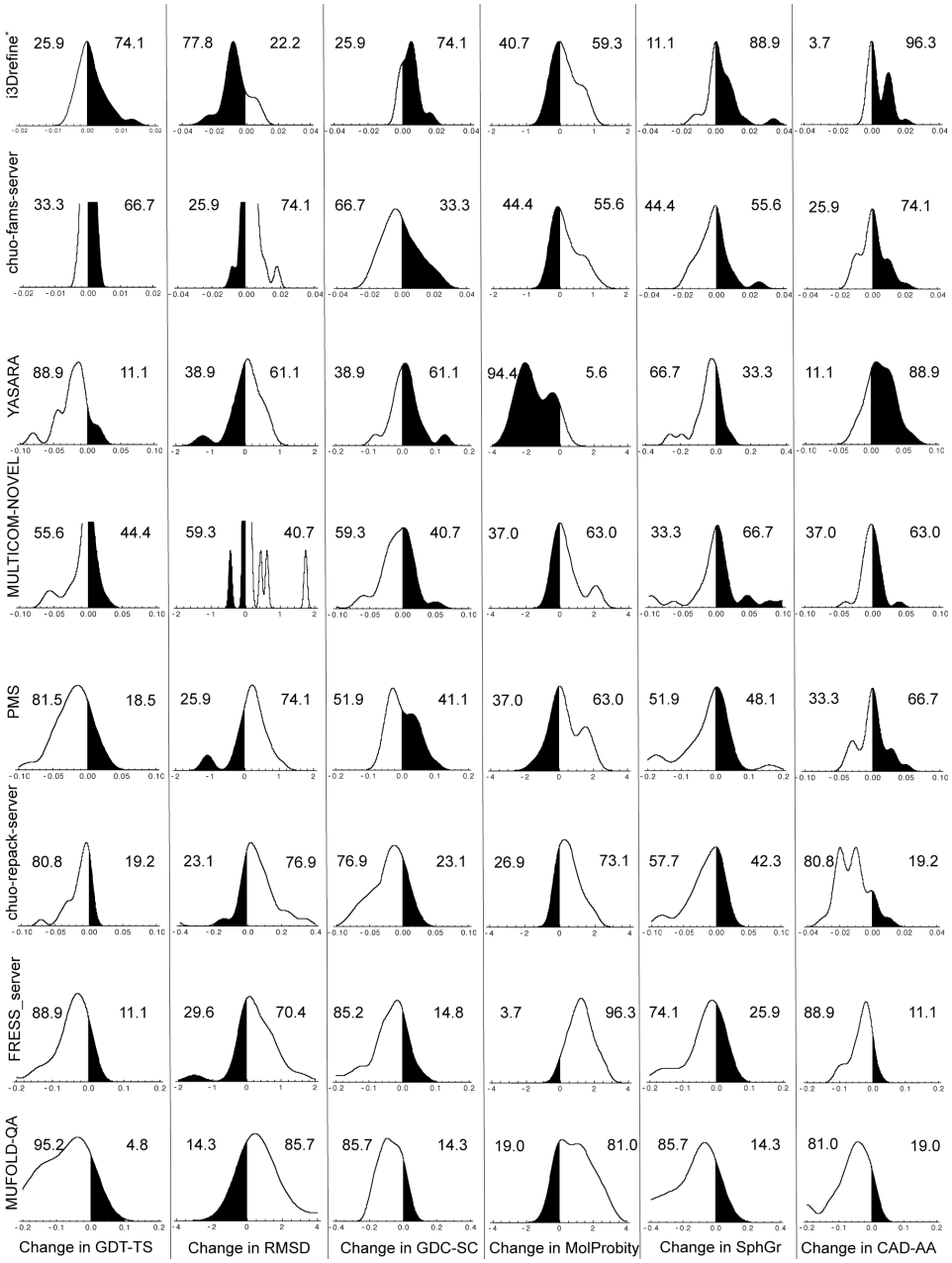


**Figure 3.4**. Distribution and degree of refinement for top server groups based on first submitted model.

Distribution and degree of score changes relative to starting models for the 8 groups based on the first submitted models. The X-axis shows changes in scores with respect to the starting model. Regions shaded in black indicate improvement over the starting model. The numeric values are the percentage of times the structures were made better or worse than the starting model for each metric. The groups are ordered by the sum of overall quality score.

* CASP10 group name for i3Drefine is MULTICOM-CONSTRUCT.

changes for each predictor and each quality measure. Some interesting variations between groups can be observed and often a trade-off exists between the extent of improvement and consistency. For example, groups like i3Drefine and chuo-fams-server perform modest but consistent improvement in almost all the scores. While the delta scores for these predictors usually lie within ~ ±4%, the distributions are skewed towards improvement. On the other hand, there exist more adventurous groups like YASARA and MULTICOM_NOVEL capable of performing larger improvements at the cost of consistency. Also, different server predictors excel at different aspects of refinement. For instance, i3Drefine improves GDT-TS and RMSD, GDC-SC, SphereGrinder and CAD-AA scores more frequently than any other groups. The ability of YASARA to improve the GDC-SC, MolProbity and CAD-AA scores in terms of degree of change and consistency is quite impressive. The most striking feature we observe is the inability of predictors to improve the backbone positioning as judged by GDT-TS and RMSD scores. i3Drefine is the only server method able to perform consistent improvement in backbone quality as measured by a simultaneous improvement in ΔGDT-TS and ΔRMSD scores. Clearly, most of the predictors are better at improving general physicality of the starting structures than at improving the backbone positioning.

Because the predictors often face difficulty in correctly ranking their submissions, the first models are often not the best submitted one. To overcome this challenge, we have recalculated the results by examining only the best structure for each group (as judged by $Q_{overall}$) for each target. In case the groups (like YASARA, chuo-fams-server and chuo-repack-server) submitted only one model as prediction, we are left with the only choice to

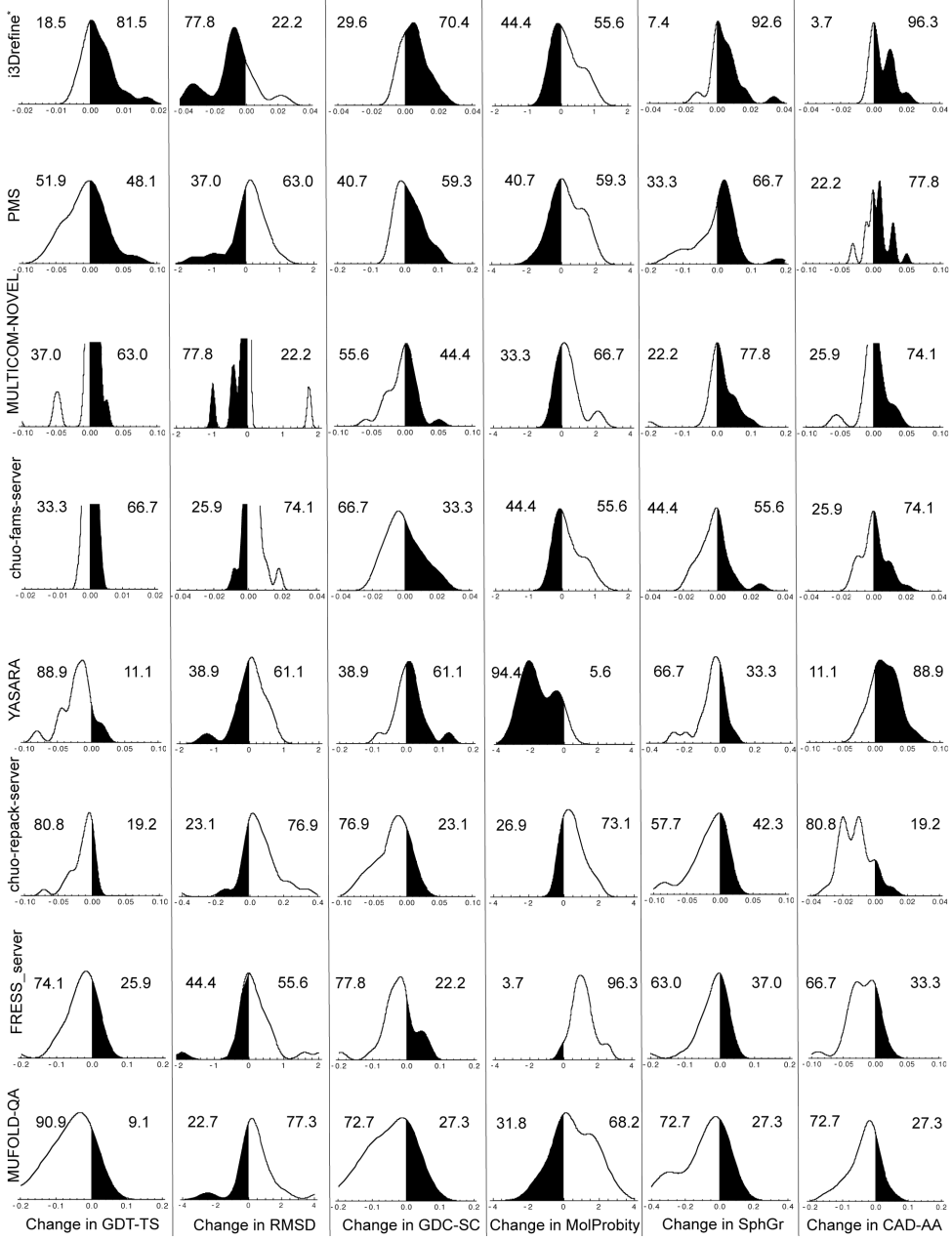


**Figure 3.5**. Distribution and degree of refinement for top server groups based on best submitted model.

Distribution and degree of score changes relative to starting models for the 8 groups based on the best submitted models as judged by quality score for each target. The X-axis shows changes in scores with respect to the starting model. Regions shaded in black indicate improvement over the starting model. The numeric values are the percentage of times the structures were made better or worse than the starting model for each metric. The groups are ordered by the sum of overall quality score.
* CASP10 group name for i3Drefine is MULTICOM-CONSTRUCT.

select that as best prediction. When judged by their best model for each target, there are two groups that perform better than the 'Void' pseudo group as shown in the lower part of Table 3.3. Once again, i3Drefine outperform all the server predictors with a consistent improvement in all the quality measures except MolProbity as evaluated by cumulative change in scores. The only other group that perform better than 'Void' is PMS with an impressive ability to improve the overall RMSD, GDC-SC and CAD-AA scores. We observe a consistent improvement in cumulative scores changes for the predictors submitting multiple predictions when the best models are selected from each group. The distributions of changes in model quality relative to the starting models are captured in Figure 3.5 with the best submitted model for each predictor as judged by six quality metrics. Once more, we see a clear trade-off between consistency and degree of refinement and different predictors performing well at different aspects of refinement. It can be observed from Figure 3.5, that i3Drefine improves GDT-TS and RMSD, GDC-SC, SphereGrinder and CAD-AA scores more frequently than any other groups, indicating its ability for a consistent improvement. The changes are, however, modest in nature. Although the group PMS has an impressive cumulative ΔRMSD score as shown in Table 3.3, Figure 3.5 reveals that changes in RMSD score is not consistent for this predictors. The overall RMSD is improved primarily due to large changes made in three targets (TR671, TR720 and TR722) and not because of consistency. When the best models are considered, MULTICOM-NOVEL has been seen to have notable ability to improve backbone positioning as measured by GDT-TS and RMSD scores by performing a consistent and often large improvement. Apart from i3Drefine, MULTICOM-NOVEL is the only other predictor able to achieve a consistent and simultaneous improvement in

ΔGDT-TS and ΔRMSD scores. YASARA group is shown to have promising ability to consistently improve MolProbity score and often with a large degree. In short, if we set aside the difficulty of the predictors to correctly rank their submissions and instead focus on the best structures from each group, we see more successes in refinement.

Overall, i3Drefine method has shown promising ability for a steady improvement in nearly all quality measures both in terms of first or best submitted predictions. The ability of i3Drefine to consistently improve GDT-TS and RMSD scores, which appear to be the most difficult metrics to improve consistently, is also encouraging.

### 3.4.5 Head-To-Head Comparison Of Server Predictors And Their Statistical Significance

Figure 3.6 shows the head-to-head comparison in the quality metrics for eight server predictors considering the first model. Upper part of Table 3.4 summarizes the *p*-values in Wilcoxon signed-rank test with null hypothesis that the refined models are same as the starting structures for eight server predictors. At 5% confidence level, i3Drefine performs statistically significant improvements in RMSD, GDC-SC, MolProbity and CAD-AA scores. The only other group with a statistically significant positive result in for at least one score is YASARA improving MolProbity score significantly. The results remain largely unaffected when judged by the best model for each target. In Figure 3.7, we present the results for eight server groups considering the overall best models for each

target, whose *p*-values of Wilcoxon signed-rank test have been shown in the lower part of Table 3.4. With the best overall model, i3Drefine performs statistically significant positive result in all quality measures except MolProbity at 5% confidence level. Strikingly, the rest of the server predictors are either indistinguishable from or worse than the 'Void' group although the magnitude of average change in scores differs for each method. Given the small number of targets, a group must perform very consistent improvement to be statistically significant with respect to 'Void' group and promisingly, i3Drefine is the only server method participating in CASP10 refinement experiment capable to achieve statistically distinguishable improvement in most of the quality metrics.

**Table 3.4.** *p*-values of score changes (Wilcoxon signed-rank test) relative to starting model for the top server groups in CASP10 refinement experiment[*]

| Selection | Group Name | $P_{GDT-TS}$ | $P_{RMSD\ (Å)}$ | $P_{GDC-SC}$ | $P_{MP}$[a] | $P_{SG}$[b] | $P_{CAD-AA}$ |
|---|---|---|---|---|---|---|---|
| First Model | MULTICOM-CONSTRUCT[c] | 0.206377 | **0.001270** | **0.000196** | 0.102297 | **0.035156** | **0.000977** |
| | chuo-fams-server | 0.705321 | 0.002984 | 0.420913 | 0.557197 | 0.075062 | 1.000000 |
| | YASARA | 0.001715 | 0.663197 | 0.338008 | **0.000232** | 0.029442 | 0.012940 |
| | MULTICOM-NOVEL | 0.101554 | 0.800817 | 0.075386 | 0.107752 | 0.982396 | 0.629059 |
| | PMS | 0.000590 | 0.046136 | 0.782327 | 0.062602 | 0.248658 | 0.930290 |
| | chuo-repack-server | 0.000123 | 0.002466 | 0.000764 | 0.000526 | 0.006444 | 0.000011 |
| | FRESS_server | 0.000024 | 0.005522 | 0.000618 | $6.64*10^{-6}$ | 0.004414 | 0.000017 |
| | MUFOLD-QA | 0.000021 | 0.000392 | 0.000447 | 0.001226 | 0.000214 | 0.000122 |
| Best Model | MULTICOM-CONSTRUCT[c] | **0.011201** | **0.004240** | **0.000108** | 0.178480 | **0.004181** | **0.000488** |
| | PMS | 0.312444 | 0.367622 | 0.121221 | 0.107443 | 0.763939 | 0.115318 |
| | MULTICOM-NOVEL | 0.555264 | 0.009355 | 0.485959 | 0.033486 | 0.146779 | 0.803619 |
| | chuo-fams-server | 0.705321 | 0.002984 | 0.420913 | 0.557197 | 0.075062 | 1.000000 |
| | YASARA | 0.001715 | 0.663197 | 0.338008 | **0.000232** | 0.029442 | 0.012940 |
| | chuo-repack-server | 0.000123 | 0.004268 | 0.000764 | 0.000526 | 0.006444 | 0.000011 |
| | FRESS_server | 0.001396 | 0.085835 | 0.014257 | $7.44*10^{-6}$ | 0.015651 | 0.000402 |
| | MUFOLD-QA | 0.000112 | 0.004061 | 0.004277 | 0.027268 | 0.003859 | 0.001248 |

[*] Numbers in bold indicates statistically significant positive results at P = 0.05.

[a] P-values for change in MolProbity score.

[b] P-values for change in SphereGrinder score.

[c] CASP10 group name for i3Drefine.

Comparison between first model and the best model of i3Drefine shows that effectiveness of the iterative version of the protocol against the non-iterative version (3Drefine). Except MolProbity, the iterative version enhances all the quality measures in terms of cumulative improvement relative to starting models as shown in Table 3. Also, the *p*-values of Wilcoxon signed-rank test are lower for the best model compared to the first model in GDT-TS, RMSD, GDC-SC, MolProbity and CAD-AA scores as reported in Table 3.4. In short, the degree of refinement as well as their statistical significance in the iterative version is, therefore, more pronounced than the non-iterative version of the protocol.

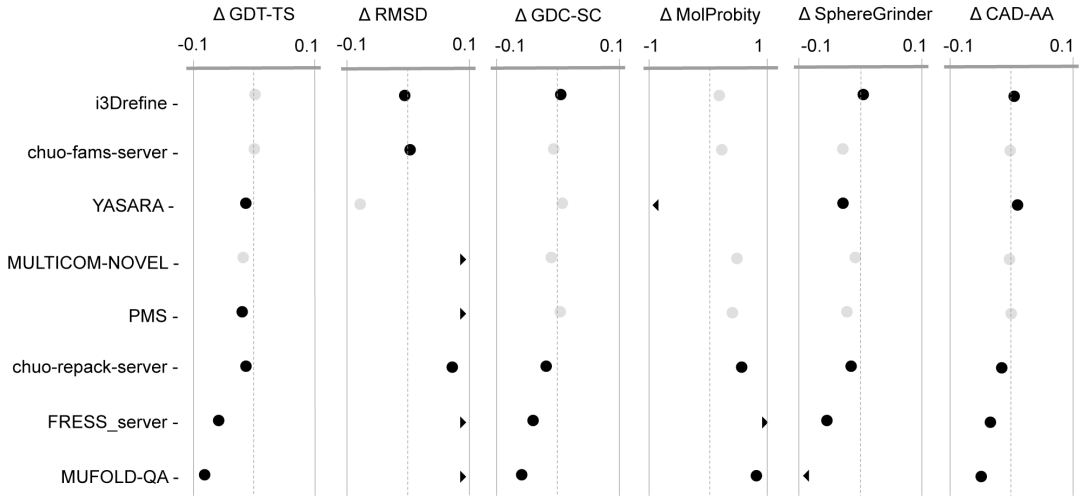


**Figure 3.6**. Summary of the average score changes and their statistical significance for top server groups based on first submitted model.

Average score changes and their statistical significance relative to starting models for the 8 groups based on the first submitted models as judged by quality score for each target. Each column shows one of the metrics we used to evaluate performance. The scales are marked at ± Average Changes relative to the 'Void' group. For GDT-TS, GDC-SC, SphereGrinder and CAD-AA scores, positive changes indicate the quality of the model has been improved by refinement whereas for RMSD and MolProbity, negative changes represent improvement. Black points are statistically distinguishable from the 'Void' group; gray points are indistinguishable (Wilcoxon signed-rank test, P = 0.05). A chevron indicates that the corresponding score is off the scale.

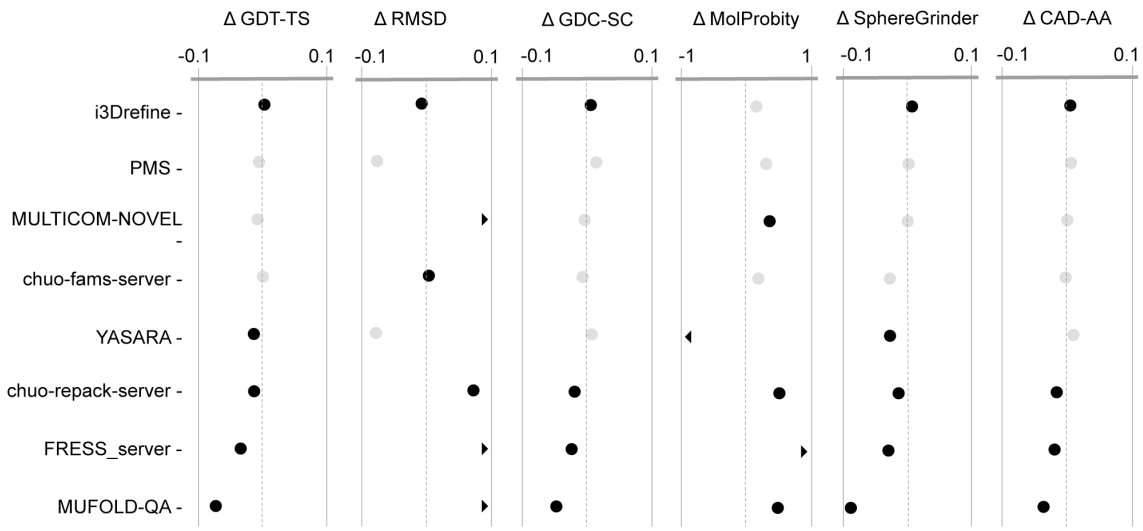


**Figure 3.7**. Summary of the average score changes and their statistical significance for top server groups based on best submitted model.

Average score changes and their statistical significance relative to starting models for the 8 groups based on the best submitted models as judged by quality score for each target. Each column shows one of the metrics we used to evaluate performance. The scales are marked at ± Average Changes relative to the 'Void' group. For GDT-TS, GDC-SC, SphereGrinder and CAD-AA scores, positive changes indicate the quality of the model has been improved by refinement whereas for RMSD and MolProbity, negative changes represent improvement. Black points are statistically distinguishable from the Null group; gray points are indistinguishable (Wilcoxon signed-rank test, P = 0.05). A chevron indicates that the corresponding score is off the scale.

### 3.4.6 Comparison Of i3Drefine With Top Five Human Predictors Participating In CASP10

Figure 3.8 shows the quartile plots of change in model quality relative to the starting model in six quality metrics for all submitted model by top five human predictors as per the official CASP10 results released during CASP10 meeting and i3Drefine for all CASP10 refinement targets. The most obvious added benefit of human predictors is the ability to perform large improvement in model quality. Groups like FEIG, Seok, Mufold and FLOUDAS seem to perform large changes in starting structures. Although the degree of refinement in these adventurous refinement strategies are much more pronounced than i3Drefine, these methods often lack the ability to perform consistent improvement. Encouragingly, the ability of i3Drefine to perform steady and consistent improvement is noticeable even when it is compared with non-server methods participating in CASP10 refinement experiment. Majority of the times, i3Drefine improves all the quality scores except MolProbity. KnowMIN protocol seems to be more conservative refinement approach than other top-performing human groups. Except SphereGrinder, KnowMIN group improves in the other quality metrics consistently. Among the top-performing human predictors, FEIG group is particularly noteworthy in its ability to improve the backbone positioning as measured by GDT-TS score accompanied by enhancement in local quality measures like MolProbity and CAD-AA. This is possibly achieved through a broader sampling around the starting model. It has to be noted, however, that the human predictors were given three weeks deadline to submit the refined structures to the prediction center as opposed to three days deadline offered for the server methods and

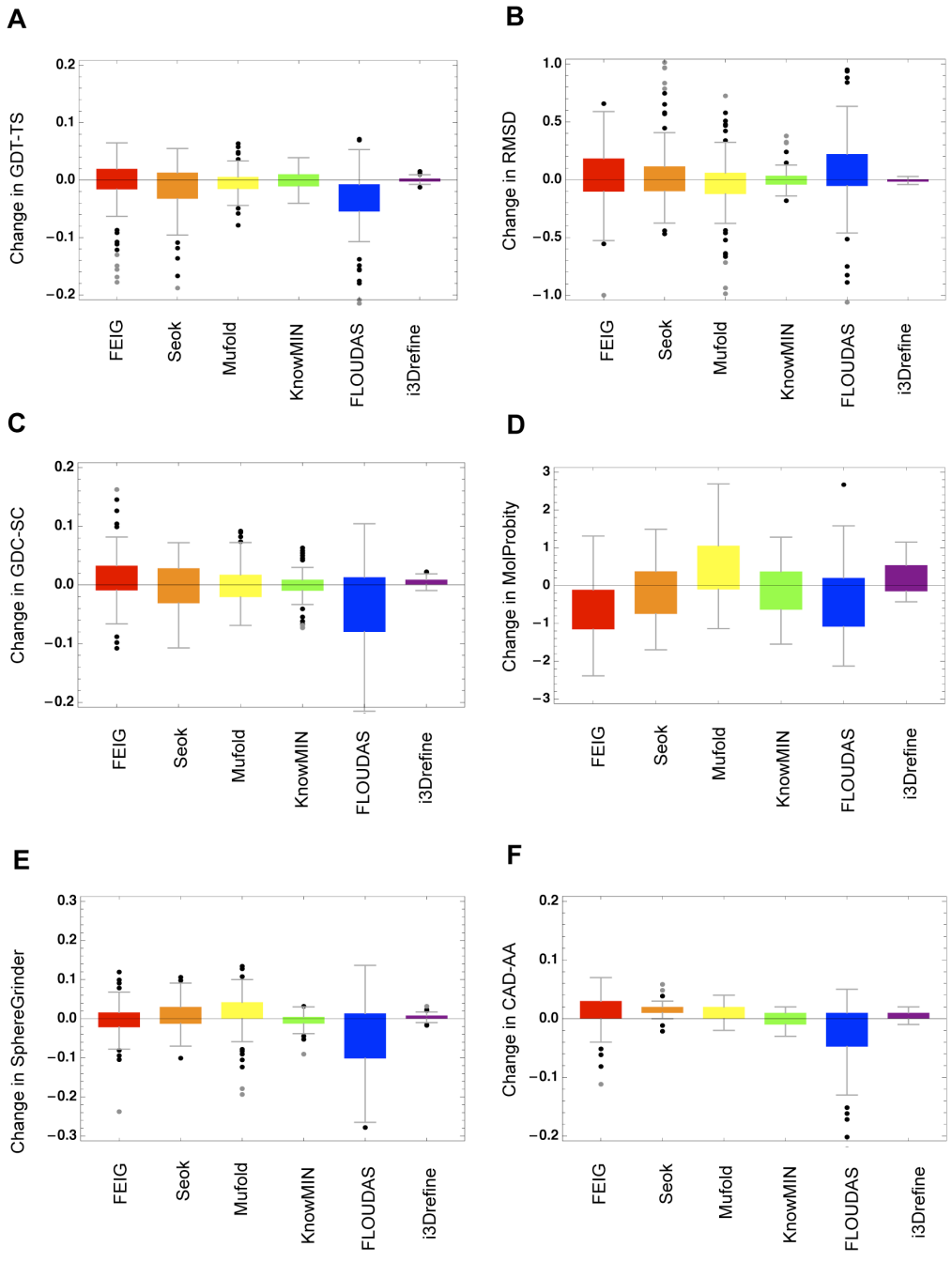


**Figure 3.8**. Quartile plots of score changes with respect to the quality of starting structures for top human predictors and i3Drefine.

Quartile plots of score changes relative to starting models for 5 human predictors and i3Drefine are shown for these metrics: (A) GDT-TS, (B) RMSD, (C) GDC-SC, (D) MolProbity, (E) SphereGrinder and (F) CAD-AA. The points outside the boxes indicate the outliers.

there might be significant human intervention involved in the non-server prediction primarily because of the relaxed submission window. A server group like MULTICOM-CONSTRUCT (i3Drefine), on the other hand, has to be completely automated in order to meet the submission deadline. It is, therefore, unfair to directly compare a server method with human groups especially when the turnaround time for a human predictor is not known. Nevertheless, the ability of human predictors to perform larger improvement can advance the field of protein structure refinement; thereby enhancing the accuracy of contemporary computational protein structure prediction methods provided these methods could be automated providing the prediction within a reasonable amount of time. In addition to being directly implemented in an automated server, human predictors in the CASP experiments often generate valuable insights and guidance for improving protein structure refinement in general.

## 3.5 Conclusions

In this work, we present a computationally inexpensive and reliable protocol for protein structure refinement, called i3Drefine and systematically analyze its performance in a completely blind mode on the targets issued for refinement category in recently concluded CASP10 experiment based on a diverse set of quality metrics. When compared with other state-of-the-art server predictors participating in CASP10, i3Drefine is observed to perform more consistently than other methods. Future directions would be to explore the possibility of i3Drefine method to perform larger improvement the quality

measures by performing a broader sampling around the starting model and possible amendments to the composite force filed. The executable version of i3Drefine software is freely available to the community providing open access to an efficient refinement method. The low computational cost and high accuracy of the i3Drefine protocol will allow this consistent refinement method to be run on a genome scale or be adopted as a final step in computational structure prediction pipeline.

# Chapter 4

# Protein Structure Refinement By Iterative Fragment Exchange

## 4.1 Abstract

Despite significant advancement of computational methods in protein structure prediction during the last decade, these techniques often cannot achieve allowable prediction accuracy to be applied in solving biological problems. Bringing these low-resolution predicted models to high-resolution structures close to their native state, called the protein structure refinement problem, however, has proven to be extremely challenging and a largely unsolved problem in the field of protein structure prediction. Here, we propose a new approach to protein structure refinement by iterative fragment exchange, called REFINEpro. The protocol first identifies the less conserved local regions in the initial model by consensus approach using ensemble of models produced for the same protein target. We call these regions problematic regions (PRs). The qualities of the PRs are then iteratively improved by exchanging better-modeled fragments corresponding to these PRs from structures in the ensemble. This method has been tested on benchmark

datasets comprising of decoys generated through both template-based and ab-initio protein structure prediction methods and exhibits promising improvement over the initial models in both global and local model quality measures, indicating a new avenue to solve the protein structure refinement problem. REFINEpro web server is freely available at http://sysbio.rnet.missouri.edu/REFINEpro/.

## 4.2 Introduction

The advancement of template based modeling (TBM) techniques and the expansion of protein sequence and structure spaces have certainly improved the average qualities of protein models over the previous decades. However, the contemporary computational protein structure prediction methods still lack the consistent accuracy needed to be successfully applied to address biological problems. Refinement of these predicted models in order to enhance the qualities thereby bringing them closer to the native state is, therefore, an integral part of protein structure prediction pipelines.

Efforts to solve the protein structure refinement problem have usually been rooted in two schools of thoughts. One is physics based methods which is governed by the thermodynamic hypothesis proposed by Anfinsen that the native structure of a protein corresponds to the global minimum of its free energy [19]. Consequently, a force-field is first developed to calculate the potential energy of the initial protein model. Then the potential energy is minimized through conformation changes with the goal to find the free-energy minimum in the protein energy landscape using traditional molecular mechanics (MM) potentials like AMBER99 [20, 21], OPLS-AA [22], etc. A number of

noteworthy studies have been performed in this direction [23, 24]. However, there are two major bottlenecks of these methods: (1) limited accuracy of physics based empirical force fields and (2) "multiple-minima problem" arising from the presence of many local minima in protein's multidimensional energy landscape [25]. The other school of thoughts is "knowledge-based" methods that utilize the statistical potentials derived from the analysis of recurrent patterns in experimentally derived protein structures and sequences [26]. Molecular Dynamics (MD) simulation is widely used in this kind of protocols [27, 28] to move every atom of a protein. Apart for some isolated cases, however, no systematic structural improvement has been attained [29].

Some promising progress has been made in the recent past by combining the two schools of thoughts, that is, by using composite physics and knowledge-based potentials [30-32] to solve protein structure refinement problem. Although encouraging, these techniques highlight a key issue in protein structure refinement – that is, majority of these methods follow a conservative local sampling strategy around the starting structures producing improvement only in local qualities of the models rather than substantially improving the backbone positioning. Development of a method capable of performing global refinement aiming to resolve differences in the overall fold of the protein model is, therefore, a crucial step forward for solving protein structure refinement problem and more generally, towards the improvement of computational protein structure prediction.

We previously developed a refinement procedure, called 3Drefine [98] by optimizing the hydrogen bonding network and atomic level energy minimization using a composite physics and knowledge-based force field. We now extend the approach by partitioning the structure refinement process into two stages: (1) global refinement with the goal of

improving the overall fold of the starting model and (2) local refinement which aims at correction of local errors like irregular hydrogen bonding, steric clashes, unphysical bond length, unrealistic bond angles, torsion angles and side-chain χ angles. These two stages demands different approaches and in accordance with the previous studies [48, 101], we decided to perform them sequentially. The first stage is accomplished by iterative fragment exchange based on an ensemble and we use our previous method 3Drefine to perform the refinement of the general physicality of the models. The new method, named REFINEpro, has been tested on diverse and independent benchmark datasets and has demonstrated significant potential in simultaneous improvement of the overall fold and resolving the local errors, thereby improving both the global and local qualities of the starting models.

## 4.3 Materials And Methods

The REFINEpro protocol is shown in Figure 4.1. Given a target protein structure and a model ensemble consisting of numerous structures generated for the same target, the method first identifies the problematic regions (PRs) in the initial model using a consensus method. An iterative refinement is then applied for each PR via generation of hybrid models by combining the initial model and the PR exchanged by structures from

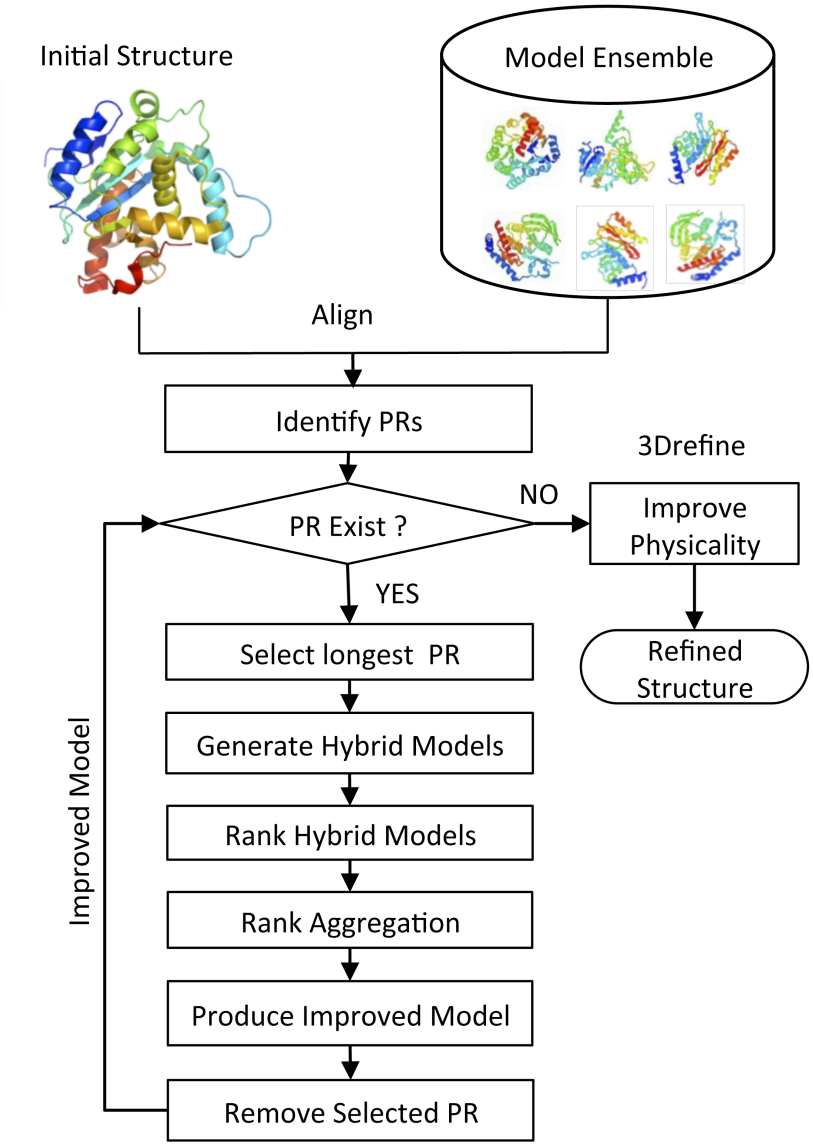


**Figure 4.1**. Flow Chart of the REFINEpro protocol. The protocol indicates stages of the iterative refinement including identification of PRs in the initial model, generation of hybrid models, quality assessment to produce the improved model followed by correcting local errors using 3Drefine to produce refined model.

the ensemble followed by quality assessment to select the best hybrid model. Finally, atomic level energy minimization is performed on the best hybrid model using 3Drefine to optimize the hydrogen-bonding network and to improve the local qualities in order to produce the refined structure. The procedure is fully automated and the average running time for each refinement target is less than 4 hours at a 2.4 GHz CPU. The REFINEpro web server is freely available at http://sysbio.rnet.missouri.edu/REFINEpro/.

### 4.3.1 Predicting PRs

It is vital to identify appropriate regions of refinement in the initial model. In template based modeling, the regions build from reliable template information tend to be largely correct and any attempt to refine them may pose the risk of degrading the model quality. The key is, therefore, to detect the problematic regions (PRs) in the starting structure and then try to refine them while keeping the conformation of the reliable regions unaltered. In REFINEpro, consensus local quality assessment technique was adapted using a model ensemble approach. For identification of PRs, we used a quality measure, similar to one originally developed by Levitt and Gerstein [88] and later widely used in developing global structural similarity measures such as TM-score [84] or MaxSub [102] and for local quality assessment of protein models [103-105]. This is called S-score.

In the first step, the models in the ensemble with missing residues were discarded and the valid structures were trimmed to exactly match the residues in the starting model. Then, each model in the ensemble were superposed into the initial model using TM-score program [84]. Once the superposed model pool was organized, we calculated the distance

between $C_\alpha$ atoms after optimal structural superposition between the initial structure and each model in the ensemble. The distance was converted into the S-score using the equation:

$$S_i^j = \frac{1}{1 + \left( \frac{d_i^j}{d_0} \right)^2} \quad (1)$$

where $S_i^j$ is the S-score of $i^{th}$ residue of the starting model with respect to the $j^{th}$ model in the ensemble, $d_i^j$ is the calculated $C_\alpha$ distance between residue i of the starting model and the corresponding residue of the $j^{th}$ model in the ensemble and $d_0$ is the distance threshold. The purpose of converting the distance into S-score is that there is no upper limit to the distance and it is, therefore, often dominated by outliers. We used $d_0 = 5$Å as in LGscore [88]. For smoothing purpose and to avoid local fluctuations, we adopted a sliding window approach. The central residue and their sequence neighbors were selected in a window of fixed size of three residues and then the average S-score was calculated in the window to determine the correctness of the central residue. Therefore, the $\mu_i^j$, the average S-score of $i^{th}$ residue with respect to the $j^{th}$ model in the ensemble is defined as:

$$\mu_i^j = \frac{1}{3} \sum_{k=i-1}^{i+1} S_k^j \quad (2)$$

We then use the decision function $p_i^j$ to determine whether residue i in the starting model have more than 5Å deviation compared to model j in ensemble as below:

$$p_i^j = \begin{cases} 1 : \mu_i^j < 0.5 \\ 0 : otherwise \end{cases} \quad (3)$$

where $p_i^j$ assumes the value 1 if the $i^{th}$ residue in the initial model deviates more than 5Å (i.e. $S_i^{win} < 0.5$) with respect to $j^{th}$ model in the ensemble and 0 otherwise. The empirical likelihood of residue conservation in a model ensemble is defined using the equation:

$$RC_i = \frac{1}{n}\sum_{k=1}^{n} p_i^k \quad (4)$$

where $RC_i$ is the probability of the $i^{th}$ residue in the initial model to have average fluctuation more than 5Å for all models in the ensemble and n is the number of valid models in ensemble. Higher $RC_i$ values indicates more fluctuation and therefore, more likely to be problematic. The rationale behind this step is: reliable residues tend to be conserved across the model ensemble, but the problematic residues prone to have higher diversity in terms of backbone $C_\alpha$ positioning. The threshold of $RC_i$ (hereafter called the residue conservation index) was chosen to be 0.5. This means if majority of structures in the ensemble have average fluctuation of more than 5Å compared to initial model (i.e. $RC_i > 0.5$), we consider the residue as problematic. Any region in the initial model having more than five consecutive problematic residues was considered as PR.

## 4.3.2 Generating Hybrid Models And Quality Assessment

Once the PRs in the initial model are identified, we attempt to refine each PR by exchanging better modeled fragments from the ensemble for the corresponding PR while keeping the reliable regions fixed. Our hypothesis is: if a better modeled fragment for a PR exists in the ensemble, then exchanging the conformation for that PR from the better quality fragment without any changes to the rest of the structure should improve the

overall fold for the PR. We adopted a hybrid model generation approach to assemble the better modeled fragments with the starting structure. First, the reliable regions were kept fixed in the initial model and the coordinates of the PR were replaced from each model in ensemble. Then, we used this model as a template for Modeller to generate hybrid model. The automodel protocol of Modeller 9v8 [106] was used to do template-based modeling with default parameter settings. We performed this operation for all the models in the ensemble to generate same number of hybrid models as the number of valid models in the ensemble with the PR replaced.

The next step is to perform quality assessment of the generated hybrid models to select structures having enhanced overall qualities compared to the initial model. To this end, six complementary single-model quality evaluation methods were applied to rank each hybrid model in the ensemble: (1) RWplus [107] which is a side-chain orientation dependent atomic statistical potential, (2) Distance-scaled, Finite Ideal-gas REference (DFIRE) [108, 109] that is based on the orientation angles involved in dipole-dipole interactions, (3) Discrete Optimized Protein Energy (DOPE) [110] which is an atomic distance-dependent statistical potential derived from a pool of known protein structures by applying probability theory, (4) FRST [111] that is based on weighted combination of four complementary knowledge-based potentials: (i) the RAPDF potential [112], (ii) solvation potential, (iii) hydrogen bond potential and (iv) torsion angle potential, (5) TAP [113] which measures the local sequence to structure fitness of the protein model depending on the torsion angle propensities, and (6)

ModelEvaluator [114] that is a machine learning approach based on features derived from secondary structure, relative solvent accessibility, contact map, and beta sheet structure.

It is worth mentioning here that along with the hybrid models, we also used the starting structure during ranking. This was performed with the vision that in case there is no fragment present in the ensemble which can improve the quality of the PR, then the initial model should remain unaltered.

### 4.3.3 Consensus Ranking Of Hybrid Models

After ranking each of the hybrid models using six above mentioned model quality assessment methods, the obvious subsequent phase is to select the top ranked model and designate

that as the improved model for the PR. But, because of the complementary nature of the evaluation methods, the ranking is often inconsistent across different protocols. One way to overcome this obstacle is to apply cumulative or average ranking in order to identify the top ranked model. However, if average rank (or cumulative rank) is applied to derive the consensus ranking, we see many ties between different hybrid models which make it difficult to identify the best model. Thus, arriving at the optimal consensus ranking for all the hybrid models in order to select the best model is a non-trivial problem. Our goal here is to find an optimal ranking which would be as close as possible to all individual ranking schemes simultaneously. This problem, therefore, can be viewed as an optimization problem.

The objective function takes the following form in its abstract representation:

$$\varphi(\delta) = \sum_{i=1}^{m} w_i d(\delta, L_i) \tag{5}$$

where $\delta$ is a proposed ordered list of length $k = |L_i|$ with m number of individual ordering, $w_i$ is the importance weight associated with list $L_i$, $d$ is a distance function described below and $L_i$ is the $i^{th}$ ordered list. The aim would be to find $\delta$* which would minimize the total distance between $\delta$* and $L_i$'s. This is denoted as:

$$\delta^* = \arg\min \sum_{i=1}^{m} w_i d(\delta, L_i) \tag{6}$$

We adopted a weighted rank aggregation method to derive the optimal solution using a R implementation [115]. Cross-Entropy Monte Carlo algorithm (CE) [116, 117] was applied using the Kendall's Tau distance for partially ordered lists [118]. The Kendall's Tau distance counts the number of pair-wise disagreements between different lists, and normalizes by the maximum possible disagreements. When the Kendall's Tau distance is 0, the two lists are exactly the same, and when it is 1, they are in reverse order. Two random lists have, on average, a distance of 0.5. The CE algorithm involves 200,000 maximum iterations or until convergence with a convergence indicator of 20 meaning that if the smallest value of the objective function does not change during 20 consecutive iterations during the optimization process, the algorithm is assumed to be converged to its optimal solution. We used the same weight ($w_i$ =1) for all the six different ranking schemes. The optimal solution derived by CE algorithm is finally chosen as the consensus ranking of the hybrid models. We term this "optimal ranking". To the best of our knowledge, weighted rank aggregation has not been used before in consensus model quality assessment. Moreover, the application of rank aggregation in the field of protein structure refinement is new.

### 4.3.4 Improving Overall Fold In The Initial Model Through Iterative Fragment Exchange

With the optimal ranking at hand, the models ranked above the starting structure were extracted to construct what we call “superior model set”. Then, a maximum of top three models were chosen from the superior model set as templates to feed into automodel class of Modeller 9v8 [106] to derive the “improved model” for the PR using multiple template alignment. It has to be noted here that selecting the number of models to be fed into Modeller depend on the structures present in the superior model set. For instance, if only two hybrid models are ranked above the starting structure after optimal ranking, then the superior model set consists of two models and we use only them as templates. In case the superior model set is empty, suggesting no hybrid model is ranked above the starting structure, the improved model is same as the initial structure.

When the starting structure consists of multiple PRs, we employed iterative refinement strategy to gradually improve the initial model with each PR getting improved in a single iteration. The PRs were sorted based on their length and longer PRs getting higher priority than shorter PRs. After each iteration, the improved model corresponding to a PR becomes the starting model for the next round of iteration aiming to improve the next PR. This process continues until all the PRs in the initial model are consumed.

### 4.3.5 Improving General Physicality To Produce The Refined Structure

When all the PRs in the initial model are iteratively refined and the final improved model is generated, the overall fold of the starting structure is supposedly improved. We now aim to improve the local errors and general physical reasonableness of the final improved model like reducing any unfavorable steric clashes or staggered $\chi$ angles. This was achieved by applying atomic level energy minimization using our previously developed refinement procedure, 3Drefine. Our previous study shows that 3Drefine has been reliable in consistent improvement of the local qualities of protein models [98]. Also, because of the fast running time of 3Drefine, it does not pose any computational overhead to the REFINEpro pipeline. The energy minimized model is the refined model.

### 4.3.6 Metrics Used For Evaluation

From the flowchart of REFINEpro (Figure 4.1), it is clear that we need two-fold evaluation method for our refinement pipeline. First, we are interested to see how accurately REFINEpro predicts the PRs in the starting models and second, how consistently our method can improve the global and local qualities of the initial structures to bring it closer to native state.

#### 4.3.6.1 Assessment Criteria For PR Prediction

In order to identify the true problematic residues in the initial model, we superposed the initial model on the native structure using TM-score and calculated the distance between $C_\alpha$ atoms after optimal structural superposition. The distance is converted to the S-score using Eq. (1). Once again, we used a sliding window of 3 residues around the central residue to avoid local fluctuation and then calculated the average S-score in that window to determine the correctness of the central residue using Eq. (2).

We use receiver operating characteristic (ROC) curve to evaluate the overall prediction accuracy of problematic residues. ROC curve is a plot of the sensitivity versus (1 - specificity) for a binary classifier as its decision boundary is moved. Sensitivity measures the capability of predicting positive samples (problematic residues in our case) correctly and specificity determines if any non-problematic residues are incorrectly predicted as problematic residues.

Since problematic residues prediction becomes a binary classification problem when the residue conservation index is set at 0.5, we measure its performance by using the following widely used criteria functions:

$$Precision = \frac{TP}{TP + FP} \qquad Specificity = \frac{TN}{TN + FN}$$

$$Recall = \frac{TP}{TP + FN} \qquad Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where true positive (TP) is the number of true problematic residues that are predicted

correctly, true negative (TN) is the number of true non-problematic residues that are predicted correctly, false positive (FP) is the number of true non-problematic residues that are predicted to be problematic and false negative (FN) is the number of true problematic residues that are predicted to be non-problematic.

### 4.3.6.2 Model quality evaluation measures

We assess the quality of the models from two perspectives: (1) similarity to the native structures and (2) physical reasonableness of the models. $C_\alpha$ Root Mean Square Deviation (RMSD) [85] is used to evaluate global positioning of $C_\alpha$ atoms purpose where a lower RMSD value indicates that the protein model is close to its native state. However, RMSD is very sensitive to small structural errors. Even if the coordinates of only a few atoms undergo large atomic changes, RMSD becomes high making it difficult to assess the overall correctness of the structure. Global quality measures like GDT-HA [83] or TM-score [84] overcomes this difficulty to a large extent. TM-score counts all the residues and tends to be more sensitive to the global topology, whereas GDT-HA count the residue pairs with distances in (0.5Å, 1Å, 2Å and 4Å), and tend to be more sensitive in capturing the errors in local fragments. Both GDT-HA and TM-score lie in [0, 1] with a higher value indicating better similarity to the native structures. GDT-HA score has been a widely used scoring function to measure the global positing of $C_\alpha$ atoms in CASP experiments [33, 36, 87]. In order to evaluate the physical reasonableness and the local errors, we use MolProbity [86] – a single and composite score to measure local model quality. MolProbity score is a log-weighted combination of the rotamer outliers, torsion-angle outliers, and steric clashes that have values outside the region of experimentally derived

standard protein structures. The MolProbity score denotes the expected resolution of the protein model with respect to standard experimental structures and therefore, lower MolProbity score indicates more physically realistic model.

## 4.3.7 Data Sets used for assessment

To benchmark the performance of REFINEpro, we collected a test set containing 163 targets: (1) 107 targets from 9$^{th}$ edition of Critical Assessment of Techniques for Protein Structure Prediction (CASP9) structure prediction category and (2) 56 targets from I-TASSER decoy set.

### 4.3.7.1 107 CASP9 targets

This dataset consists of 107 CASP9 TS targets taken from http://predictioncenter.org/download_area/CASP9/. We used the first models submitted to CASP9 TS category by our structure prediction method, MULTICOM-CONSTRUCT [82] as the initial model for each of these targets. The complete archive of submitted models by all servers had been used as the ensemble. The CASP9 dataset in most interesting in terms of practical applications: (1) it contains the best models submitted by the best research groups around the world and therefore represents the state-of-the-art in the field of protein structure prediction; (2) the high diversity of targets both in terms of length and complexity reduces bias in testing our protocol and (3) because of the popularity of TBM methods amongst the CASP predictors, REFINEpro can be evaluated in its ability to refine models produced by TBM techniques.

#### 4.3.7.2 I-TASSER decoy set

Models in this dataset contain 56 non-homologous small proteins with lengths from 47 to 118 residues. I-TASSER ab-initio modeling [119] were used to generate the backbone structure and 12,500 - 32,000 conformations were selected from the trajectories of 3 lowest-temperature replicas of the simulations. Then, iterative structure clustering [120] followed by energy minimization was performed on the selected decoys using GROMACS 4.0 simulation package [121] with OPLS-AA force field [122] for improving local qualities while keeping the topologies unchanged. The decoy set is available at http://zhanglab.ccmb.med.umich.edu/decoys. For each target, the best structure (having lowest RMSD to native) was used as the initial model while the rest of the decoy set serves as model ensemble during REFINEpro run.

## 4.4 Results And Discussion

We begin by evaluating the performance of REFINEpro to detect the PRs in the initial model for all our datasets. Then, the overall improvement in global and local qualities of the initial models generated by MULTICOM-CONSTRUCT [82] for 107 CASP9 targets are presented. Finally, we judge how significant the refinement is when the initial models and the ensembles are generated I-TASSER ab-initio simulation.

### 4.4.1 Accuracy Of PR Prediction

In Figure 4.2, the ROC curves are shown for each of the two datasets we used with the corresponding area under the curves (AUC). The values of the criteria functions are presented in Table 4.1 at residue conservation index threshold of 0.5. It can be observed that, the dataset of 107 CASP9 targets yields better performance while prediction accuracy is less significant in I-TASSER decoy set. This is mainly because in I-TASSER decoy set, the initial models and the ensembles are generated by the same prediction pipeline, and thus, the conformations of the decoys are largely same, making it difficult to successfully apply consensus prediction. These results, therefore, demonstrate that our hypothesis of residue conservation works best when complementary structures are present in the ensemble.

**Table 4.1**. Accuracy measures for the prediction of PRs at residue conservation index threshold of 0.5

| Dataset | Precision | Specificity | Recall | Accuracy |
|---|---|---|---|---|
| 107 CASP9 Targets | 85.51 | 94.14 | 71.2 | 86.63 |
| I-TASSER Decoy Set | 41.02 | 86.22 | 64.7 | 83.43 |

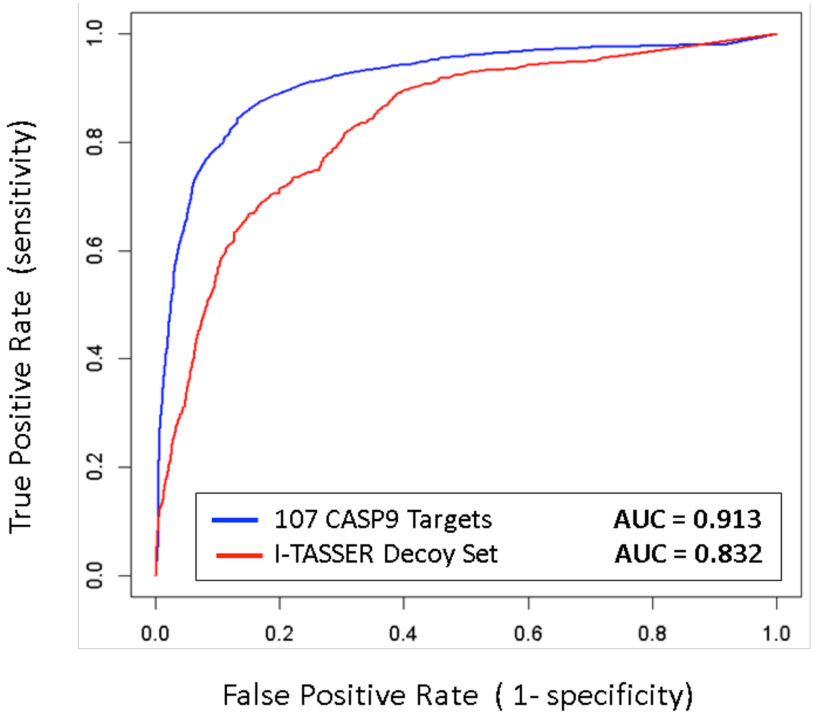


**Figure 4.2**. ROC curves for prediction of PRs.
107 CASP9 Targets (red) and I-TASSER Decoy Set (blue).
The numbers beside each legend represent the values of
Area Under the Curve (AUC).

## 4.4.2 Performance On 107 CASP9 Targets

Due to the presence of potentially disordered regions in the native structures, often there are mismatches in the CASP9 sequence with that of the experimental structures. After executing REFINEpro in blind mode, we identified the residues in the target sequences that did not have coordinates in the experimental structures by performing alignment between CASP9 sequence and the corresponding sequences for the native structures

using ClustalW [123]. These residues were removed from both the initial and refined models during refinement assessment.

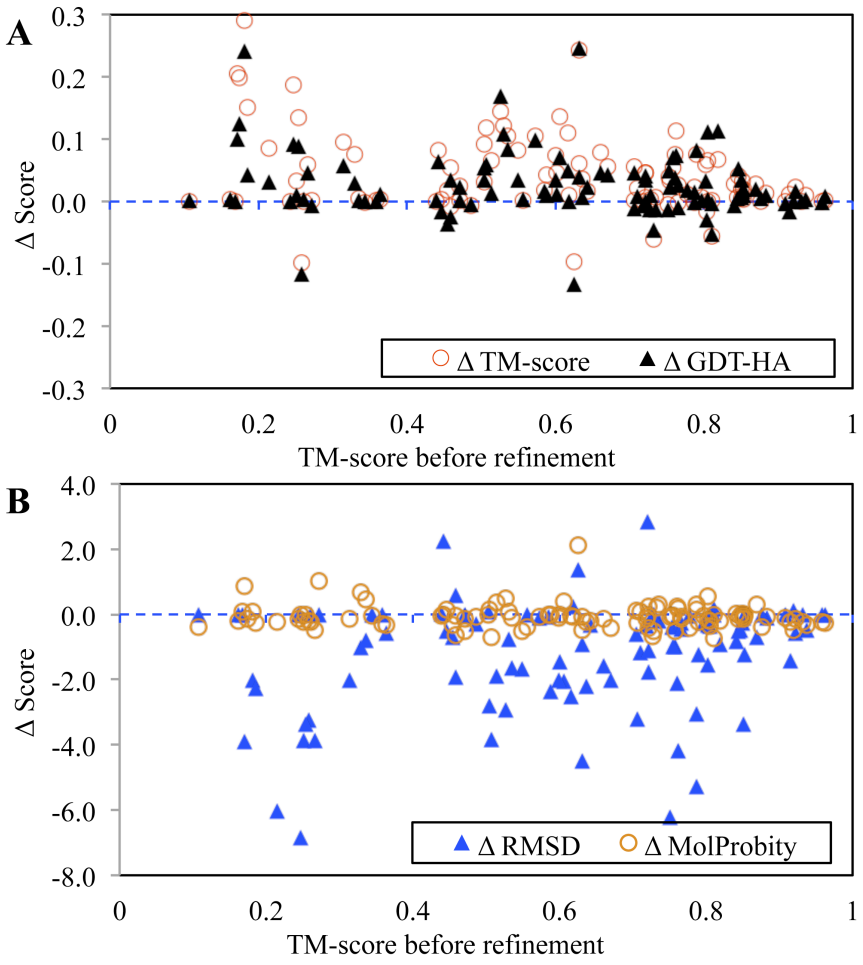


**Figure 4.3**. Changes in structural qualities using REFINEpro on 107 CASP9 targets.
(A) Scatter plot of changes in and TM-score and GDT-HA. A positive change indicates the quality of the model has been improved by refinement.
(B) Scatter plot of changes in RMSD and MolProbity score. A negative change indicates the quality of the model been improved by refinement.

Figures 4.3A and 4.3B show the scatter plot of GDT-HA, TM-score RMSD and MolProbity score difference before and after refinement against the initial TM-scores. Out of total 107 targets, REFINEpro refinement resulted in improving the model qualities for 95, 91, 78 and 80 cases with respect to RMSD, TM- score, GDT-HA and MolProbity scores respectively. Overall, 5.1% and 4.9% improvement in cumulative TM-score and cumulative GDT-HA score respectively has been observed while the average RMSD and average MolProbity improvement is 12.7% and 5.8% respectively. These results clearly demonstrate the promising ability of REFINEpro to improve the overall fold of the starting models together with enhancement in the general physicality of the models in a large benchmark set comprised of different lengths and target complexities.

A representative example of refinement is presented in Figure 4.4 for target T0559. The initial model has an RMSD of 6.188 Å with a large deviation in the N-terminal helix region compared to the native structure. After refinement, the RMSD is drastically improved to 1.694 Å with 51.1%, 38.5% and 3.6% improvement in GDT-HA, TM-score and MolProbity score respectively. The definite improvement in the N-terminal region is obvious even by simple visual inspection.

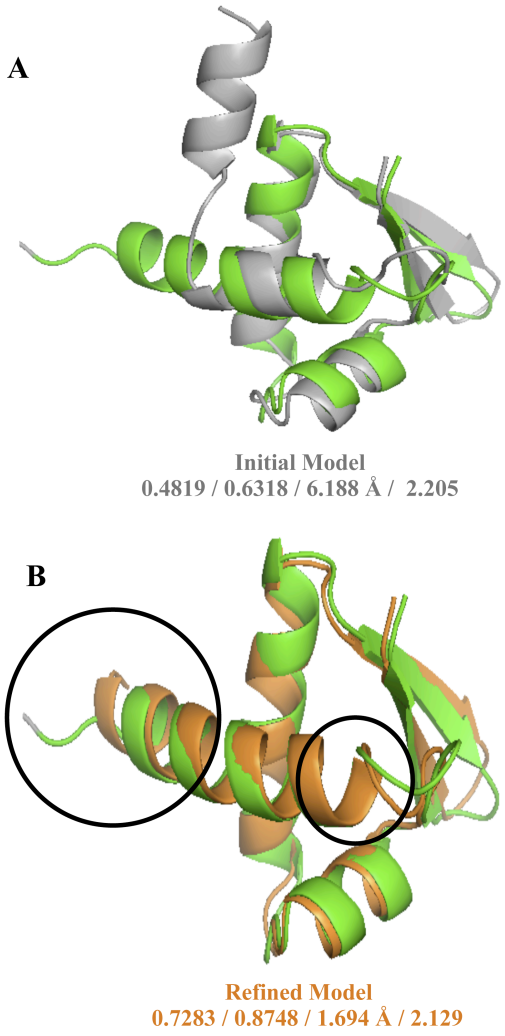


**Figure 4.4**. Example of REFINEpro refinement for CASP9 target T0559.
(A) Structural superposition of initial model (grey) on native structure (green). The values under the model indicate GDT-HA, TM-score, RMSD and MolProbity score respectively before refinement.
(B) Structural superposition of refined model using REFINEpro (orange) on native structure (green). The values under the model indicate GDT-HA, TM-score, RMSD and MolProbity score respectively after refinement. The black circles highlight the regions with prominent structural improvements.

### 4.4.3 Performance on I-TASSER decoy set

Similar to the CASP9 dataset, we performed the refinement on the I-TASSER decoy set, where initial models are generated by I-TASSER ab-initio simulation in a strict blind mode, that is, without the knowledge of the native structure.

A consistent improvement is observed in qualities of the starting structures as measured by the GDT-HA, TM-score, and RMSD scores. There were 35, 37 and 31 cases when REFINEpro brings the starting models closer to the native ones. In Figure 4.5A and 4.5B, we present the scatter plot of GDT-HA, TM-score and RMSD score difference before and after refinement against the initial TM-score for all the 56 targets. Although encouraging, the refinement for the I-TASSER decoy set is not as pronounced as the CASP9 dataset. This is primarily because the starting models in I-TASSER decoy set are already the best models selected from the ensemble with majority of the initial models have RMSD less than 3Å compared to the native structures, resulting in less PRs being identified by REFINEpro, thereby hindering the ability of REFINEpro for drastic improvement in the backbone positioning. These results, therefore, suggest that most prominent improvement in model qualities are observed in REFINEpro when the starting structure is further away from native state.

A typical example of refinement from the I-TASSER decoy set has been shown in Figure 4.6A and 4.6B for the target 1cqkA. The starting structure is quite accurate with initial RMSD of 1.448Å. The refinement is distributed across the whole sequence with reorientation of several loops to beta-strands, thereby bringing the model closer to the native state. The RMSD of the refined model is improved to 1.299Å with a 5.72%

increase in GDT-HA, 2.38% increase in TM-score and 47.2% improvement in MolProbity score.

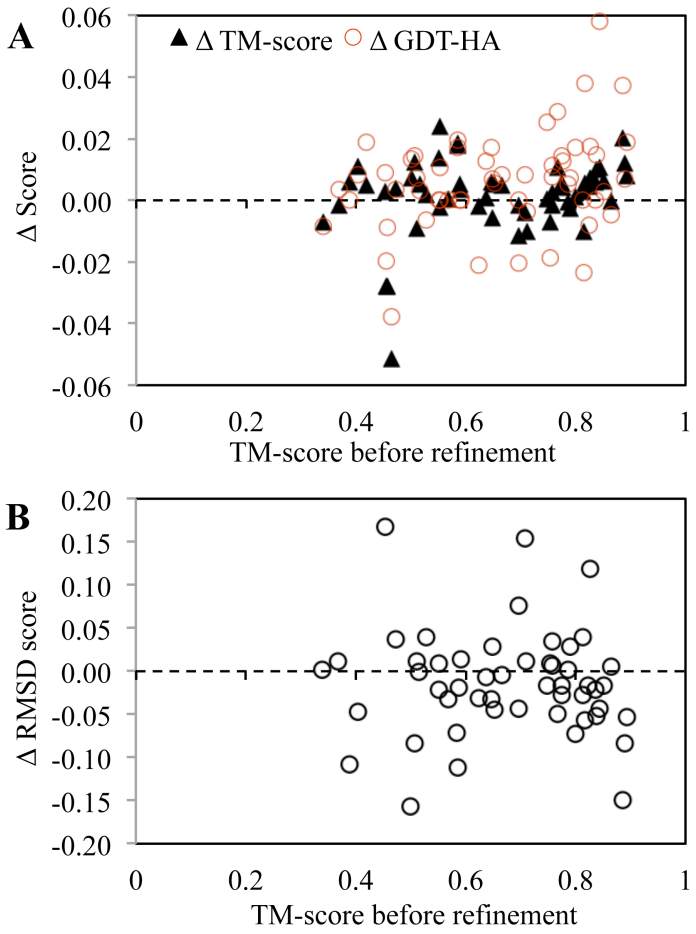


**Figure 4.5**. Changes in structural qualities using REFINEpro on I-TASSER decoy set.
(A) Scatter plot of changes in GDT-HA and TM-score. A positive change indicates the quality of the model has been improved by refinement.
(B) Scatter plot of changes in RMSD score. A negative change indicates the quality of the model has been improved by refinement.

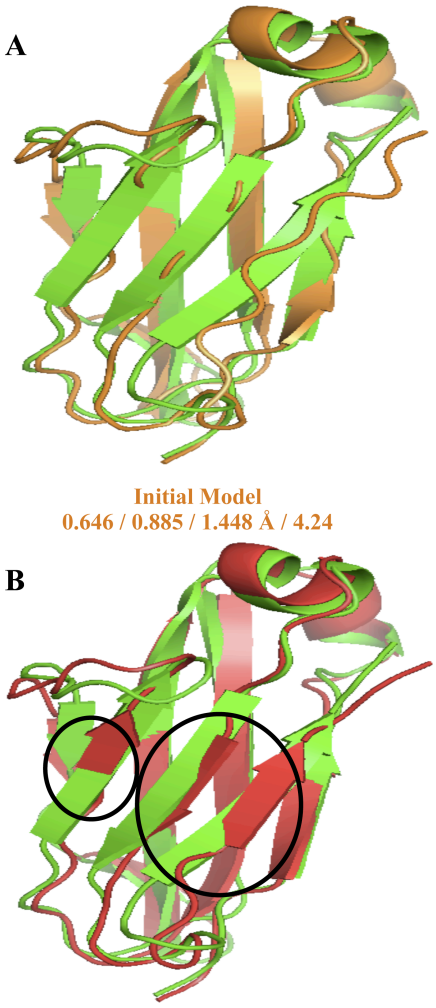


**Figure 4.6**. Example of REFINEpro refinement for I-TASSER target 1cqkA.
(A) Structural superposition of initial model (orange) on native structure (green). The values under the model indicate GDT-HA, TM-score, RMSD and MolProbity score respectively before refinement.
(B) Structural superposition of refined model using REFINEpro (red) on native structure (green). The values under the model indicate GDT-HA, TM-score, RMSD and MolProbity score respectively after refinement. The black circles highlight the regions with prominent structural improvements.

## 4.5 Conclusion

Development of a method capable of improving the overall fold in predicted protein models has been a major challenge in the protein structure refinement field. The existing state-of-the-art refinement algorithms often rely on "conservative" strategies to sample locally around the starting structures producing improvement only in the physicality of the models as opposed to improvement of the global positioning of the backbone atoms [62]. In this article, we presented a new conformation ensemble-based iterative refinement method aimed at resolving this bottleneck. Coupled with our previous study on protein structure refinement [98] the method can often drastically improve the overall fold of the initial models through refinement of loop and terminal regions or rearrangements of disoriented secondary structure segments, accompanied by correction of local errors. By performing a large-scale benchmark study on 163 targets, we demonstrated that the protocol is capable of simultaneous improvement in global and local qualities of protein models generated by both TBM and ab-initio methods. More prominent results were achieved when the model ensemble for the target structure contains diverse and complementary alternative models. To our knowledge, a fully automated ensemble based approach has not been used before in refinement problem. We hope the promising aspects of our refinement protocol provide useful insights for advancement in the field of protein structure refinement, thereby enhancing the accuracy of contemporary computational protein structure prediction methods.

Even though encouraging success has been obtained in the present study, there is still large room for improvement. The major challenges encountered were: (1) accurate

prediction of PRs in the starting structures, (2) availability of diverse and complementary models in the ensemble and (3) quality assessment method aimed at selecting the "best" hybrid model. Future directions would be to extend our consensus-based method of PR prediction to more a robust and precise approach, preferably by applying machine learning techniques. Also, significant success of REFINEpro in CASP9 dataset compared to the I-TASSER decoy set demonstrates that it is essential to have structures in the ensemble with independent and various folds. We need to investigate in future on how to automatically generate a large pool of models with various folds in a practical and efficient manner. Finally, a better model quality assessment technique is desirable which can select the "best" alternative structure from the hybrid model pool, and consequently, the accuracy of our refinement method can be improved further.

# Appendix A

## Supplementary Information for Chapter 2

This chapter provides some supplementary information related to Chapter 2.

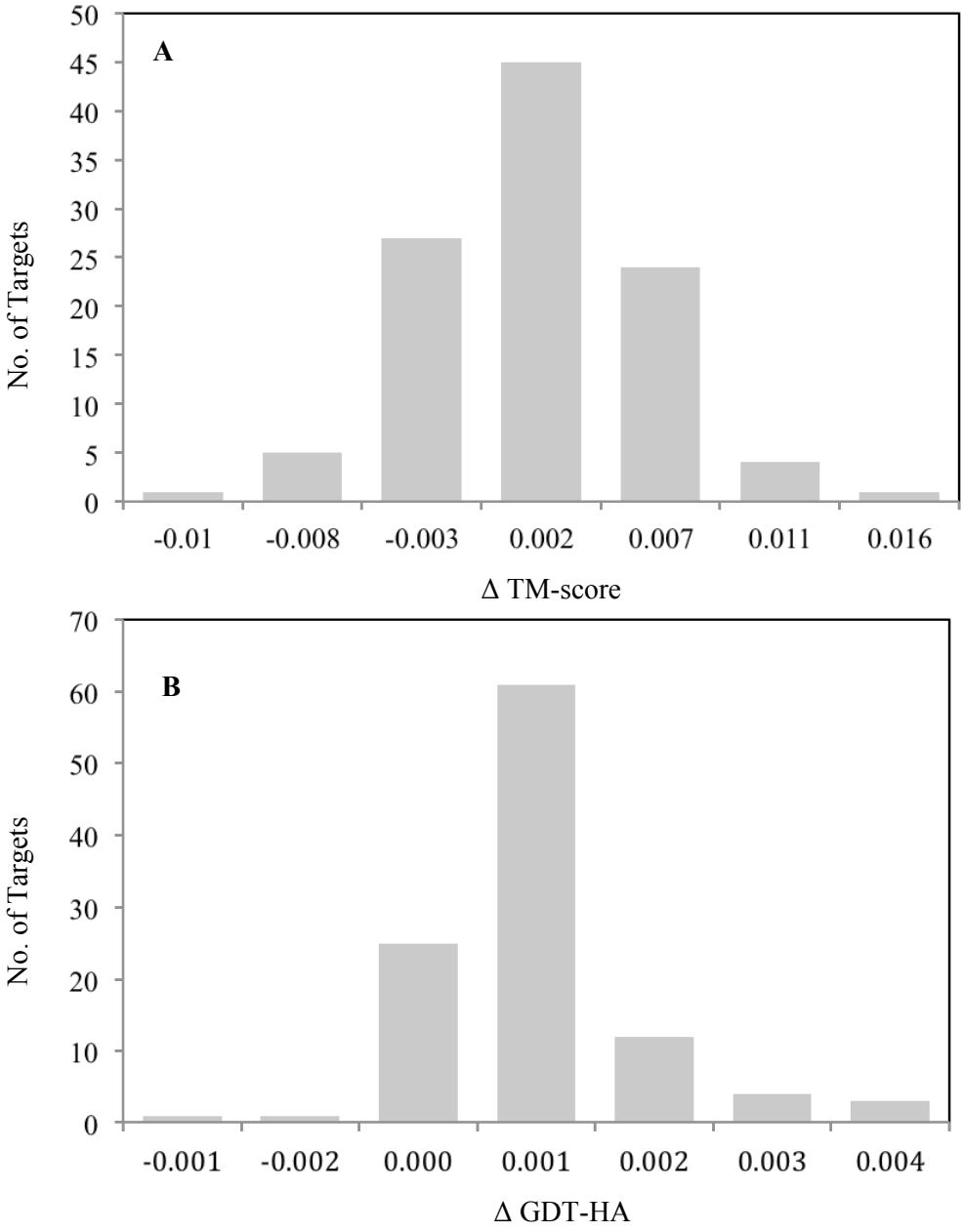


**Figure A1**. The histogram of score improvements for 107 CASP9 targets (Initial structures generated using MULTICOM-REFINE).
(A) Histogram for TM-score improvement.
(B) Histogram for GDT-HA score improvement.

**Table A1**. Refinement results for all the participating groups in CASP8 Refinement Experiment.

| Group Name | No. of Targets[a] | GDT-HA [b] | TM-score [c] | MolProbity[d] |
|---|---|---|---|---|
| 3Drefine | 12 | 6.932 | 9.3289 | 2.349 |
| NULL[e] | 12 | 6.898 | 9.316 | 2.706 |
| LEE | 12 | 6.86 | 9.195 | 2.613 |
| LevittGroup | 12 | 6.701 | 9.16 | 2.875 |
| FAMSD | 12 | 6.562 | 8.746 | 2.796 |
| SAM-T08-human | 12 | 6.523 | 9.084 | 2.762 |
| YASARARefine | 12 | 6.407 | 9.155 | 1.071 |
| Bates_BMM | 12 | 6.167 | 8.734 | 2.737 |
| FEIG_REFINE | 11 | 6.158 | 8.462 | 2.192 |
| FAMS-multi | 12 | 6.055 | 8.281 | 3.42 |
| Abagyan | 10 | 6.054 | 8.262 | 3.136 |
| xianmingpan | 11 | 5.977 | 8.474 | 1.698 |
| A-TASSER | 11 | 5.751 | 8.133 | 2.597 |
| POISE | 12 | 5.556 | 8.376 | 1.844 |
| SAMUDRALA | 10 | 5.441 | 7.19 | 2.329 |
| DBAKER | 9 | 5.413 | 7.041 | 1.636 |
| MidwayFolding | 10 | 5.273 | 7.251 | 3.528 |
| Jones-UCL | 12 | 5.175 | 8.089 | 3.402 |
| tripos_08 | 8 | 4.619 | 6.295 | 2.826 |
| Elofsson | 12 | 4.596 | 6.763 | 2.713 |
| jacobson | 7 | 4.288 | 5.566 | 2.855 |
| Keasar | 7 | 3.552 | 5.413 | 2.765 |
| POEM | 8 | 2.652 | 4.356 | 4.304 |
| PS2-server | 3 | 1.646 | 2.206 | 3.035 |
| EB_AMU_Physics | 3 | 1.61 | 2.421 | 2.867 |
| TASSER | 2 | 1.131 | 1.652 | 2.913 |
| Kolinski | 1 | 0.565 | 0.845 | 2.602 |

[a] Number of CASP8 targets in the Refinement Experiment.
[b] Cumulative GDT-HA score based on the first submitted model.
[c] Cumulative TM-score based on the first submitted model.
[d] Average MolProbity score based on the first submitted model.
[e] The initial models for the CASP8 refinement experiment.

**Table A2**. Refinement results for all the participating groups in CASP9 Refinement Experiment.

| Group Name | No. of Targets[a] | GDT-HA [b] | TM-score [c] | MolProbity[d] |
|---|---|---|---|---|
| 3Drefine | 14 | 7.388 | 10.388 | 2.101 |
| ZHANG | 14 | 7.365 | 10.396 | 3.042 |
| SEOK | 14 | 7.359 | 10.399 | 3.436 |
| NULL[e] | 14 | 7.319 | 10.368 | 2.521 |
| FAMSD | 14 | 7.284 | 10.348 | 2.55 |
| FAMS-MULTI | 14 | 7.284 | 10.348 | 2.55 |
| KNOWMIN | 14 | 7.194 | 10.182 | 2.179 |
| TASSER | 14 | 7.164 | 10.259 | 3.16 |
| BAKER | 14 | 7.156 | 10.287 | 1.327 |
| SAMUDRALA | 14 | 7.031 | 10.038 | 2.395 |
| YASARA | 14 | 7.023 | 10.043 | 0.969 |
| GWS | 14 | 6.962 | 9.74 | 2.585 |
| LEE | 14 | 6.962 | 9.74 | 2.751 |
| PCOMB | 14 | 6.876 | 9.866 | 1.407 |
| GENESILICO | 14 | 6.849 | 9.994 | 3.633 |
| SHORTLE | 14 | 6.834 | 10.04 | 3.513 |
| RECOMBINEIT | 14 | 6.779 | 9.854 | 3.177 |
| PCONS | 14 | 6.631 | 9.678 | 1.495 |
| FEIG | 13 | 6.612 | 9.26 | 2.238 |
| PROQ2 | 14 | 6.496 | 9.61 | 1.623 |
| PCONSM | 14 | 6.491 | 9.57 | 1.343 |
| PROQ | 14 | 6.246 | 9.454 | 1.542 |

[a] Number of CASP9 targets in the Refinement Experiment.
[b] Cumulative GDT-HA score based on the first submitted model.
[c] Cumulative TM-score based on the first submitted model.
[d] Average MolProbity score based on the first submitted model.
[e] The initial models for the CASP9 refinement experiment.

# Appendix B

# Software And Web-services

In this chapter, we provide a brief overview of the freely available software and web-services developed based on the aforementioned methods and protocols for the scientific community. In particular, we developed a freely available web-server named 3Drefine based on the methods presented in Chapter 2. An iterative standalone version of 3Drefine protocol, named i3Drefine, as discussed in Chapter 3, is also offered as freely downloadable software package. Finally, a publicly accessible web-service named REFINEpro based on the iterative fragment exchange method presented in Chapter 4 is made available.

## B.1 3Drefine

### B.1.1 Overview

3Drefine is a web service for consistent and computationally efficient protein structure refinement. The protocol is based on two steps of refinement process. The first step is based on optimization of hydrogen bonding (HB) network and the second step applies

atomic-level energy minimization on the optimized model using a composite physics and knowledge-based force fields. The goal of 3Drefine server is simultaneous improvement in both global and local structural qualities of the initial models to bring it closer to the native state in a computationally inexpensive manner.

## B.1.2 Availability

http://sysbio.rnet.missouri.edu/3Drefine/

## B.1.3 Input

The input of 3Drefine server must be protein structural file in PDB format containing only one chain. When the user provides initial protein structure for refinement, the server first validates the correctness of the file type. If successful, the job is queued. Otherwise, the user is shown an error message. The user needs to ensure that the initial structures in PDB format only for a successful 3Drefine run.

## B.1.4 Output

3Drefine server automatically redirects the user about the status of the current submission. After a job is completed, the refinement results can be viewed in the results page containing the initial and refined structures along with job statistics. The user can download the individual structures by clicking on them. A complete compressed archive of the job is also made available. The user can bookmark the results page to view the

results later. Alternatively, the user can search the results by specifying the Job Name and Job ID of their earlier submissions. In case the user has provided an email address, the refined structure along with the complete run achieve will be emailed immediately after the job is complete.

### B.1.5 Software Architecture

The core of 3Drefine is developed in Java (http://www.java.com/en/) on top of MESHI [65] software package. 3Drefine web server uses Perl CGI, JavaScript and PHP for a seamless user experience.

## B.2 i3Drefine

### B.2.1 Overview

i3Drefine is software for reliable and computationally efficient protein structure refinement. The goal of i3Drefine is to perform consistent and simultaneous improvement in both global and local structural qualities of the initial models to bring it closer to the native state in a computationally inexpensive manner. The name i3Drefine stands for iterative three-dimensional (3D) protein structure refinement.

### B.2.2 Availability

http://protein.rnet.missouri.edu/i3drefine/

## B.2.3 Input

i3Drefine requires only two parameters: the starting structure in PDB format and the number of refined models one wish to generate from the iterative refinement framework. Although there are no restrictions to the number of refined models one can generate, we recommend to this number within 10 so that the energy minimization does not overshoot the global minima in the multidimensional energy landscape.

## B.2.5 Output

While executing, i3Drefine automatically creates a workspace in the directory from which the program is called. This is associated with the auto-generated Job-Id of the submitted job. At the end of successful execution, RESULT/ directory created under the workspace contains the refined models numbered as REFINED_n.pdb; where n is the iteration number.

## B.2.5 Software Architecture

The core of i3Drefine is developed in Java (http://www.java.com/en/) on top of MESHI [65] software package and the command-line interface to perform the refinement is developed in Perl programming language (http://www.perl.org/). For a seamless installation and usage of i3Drefine, a Java version 6.0 or above and Perl version 5.8.8 or above is recommended. Also, since some of the energy terms in the customized force

fields require the secondary structure assignment of the starting structure for accurate calculations, DSSP program [77] needs to be used in conjunction with i3Drefine. The detailed installation instructions along with typical example of using i3Drefine have been provided in the user manual file supplied with the software. i3Drefine has been tested on 64-bit Linux based platform. However, because of the platform independent nature of Java and versatile platform support of Perl, it can be fairly easily modified to run for Windows or Mac OSX platforms.

## B.3 REFINEpro

### B.3.1 Overview

REFINEpro server is a web service for protein structure refinement using a conformation ensemble approach. The protocol is based a two stage process. In the first stage, the problematic regions in the initial protein model are identified. Then the uncertain problematic regions are replaced recursively from each of the models in the model ensemble and the best model is declared as winner for the next round of refinement. The goal of REFINEpro server is to improve structural qualities of the initial models by combining better-modeled regions from other models to bring it closer to the native state.

### B.3.2 Availability

http://sysbio.rnet.missouri.edu/REFINEpro/

## B.3.3 Input

The user needs to provide an initial protein structure for refinement along with an archive file in zip or tar.gz format containing multiple decoys for the same protein generated by same or different structure prediction methods. The server first validates the correctness of the file type. If successful, the job is queued. Otherwise, the user is shown an error message. The user needs to ensure that the structures are in PDB format only for a successful REFINEpro run.

## B.3.5 Output

REFINEpro server automatically redirects the user about the status of the current submission. After a job is completed, the refinement results can be viewed in the results page containing the initial and refined structures along with job statistics. The user can download the individual structures by clicking on them. A complete compressed archive of the job is also made available. The user can bookmark the results page to view the results later. Alternatively, the user can search the results by specifying the Job Name and Job ID of their earlier submissions. In case the user has provided an email address, the refined structure along with the complete run achieve will be emailed immediately after the job is complete.

### B.3.5 Software Architecture

The core of REFINEpro is developed in Perl programming language (http://www.perl.org/) together with bash scripting. REFINEpro web server uses Perl CGI, JavaScript and PHP for a seamless user experience.

# BIBLIOGRAPHY


[1] K. Ginalski, N. V. Grishin, A. Godzik, and L. Rychlewski, "Practical lessons from protein structure prediction," *Nucleic Acids Research,* vol. 33, no. 6, pp. 1874-1891, 2005.

[2] Y. Zhang, "I-TASSER server for protein 3D structure prediction," *BMC bioinformatics,* vol. 9, no. 1, pp. 40, 2008.

[3] A. Sali, "Modeling the structures of proteins and macromolecular assemblies," *Faseb J,* vol. 20, pp. A851-A851, 2006.

[4] P. Bradley, K. M. Misura, and D. Baker, "Toward high-resolution de novo structure prediction for small proteins," *Science,* vol. 309, no. 5742, pp. 1868-1871, 2005.

[5] A. Šali, and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints," *Journal of molecular biology,* vol. 234, no. 3, pp. 779-815, 1993.

[6] M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. Melo, and A. Šali, "Comparative protein structure modeling of genes and genomes," *Annual review of biophysics and biomolecular structure,* vol. 29, no. 1, pp. 291-325, 2000.

[7] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, "The protein data bank," *European Journal of Biochemistry,* vol. 80, no. 2, pp. 319-324, 1977.

[8] D. Baker, and A. Sali, "Protein structure prediction and structural genomics," *Science,* vol. 294, no. 5540, pp. 93-96, 2001.

[9] M. Gerstein, A. Edwards, C. Arrowsmith, and G. Montelione, "Structural genomics: current progress," *Science,* vol. 299, no. 5613, pp. 1663-1663, 2003.

[10] J. Westbrook, Z. Feng, L. Chen, H. Yang, and H. M. Berman, "The protein data bank and structural genomics," *Nucleic acids research,* vol. 31, no. 1, pp. 489-491, 2003.

[11] G. T. Montelione, "Structural genomics: an approach to the protein folding problem," *Proceedings of the National Academy of Sciences,* vol. 98, no. 24, pp. 13488-13489, 2001.

[12] D. Vitkup, E. Melamud, J. Moult, and C. Sander, "Completeness in structural genomics," *Nature Structural & Molecular Biology,* vol. 8, no. 6, pp. 559-566, 2001.

[13] Y. Zhang, and J. Skolnick, "The protein structure prediction problem could be solved using the current PDB library," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102, no. 4, pp. 1029-1034, 2005.

[14] K. Ginalski, "Comparative modeling for protein structure prediction," *Current opinion in structural biology,* vol. 16, no. 2, pp. 172-177, 2006.

[15] Z. Xiang, "Advances in homology protein structure modeling," *Current protein & peptide science,* vol. 7, no. 3, pp. 217, 2006.

[16] E. Krieger, G. Koraimann, and G. Vriend, "Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field," *Proteins: Structure, Function, and Bioinformatics,* vol. 47, no. 3, pp. 393-402, 2002.
[17] H. Wieman, E. Anderssen, and F. Drablos, "Homology-based modelling of targets for rational drug design," *Mini reviews in medicinal chemistry,* vol. 4, no. 7, pp. 793-804, 2004.
[18] R. Bonneau, J. Tsai, I. Ruczinski, and D. Baker, "Functional Inferences from Blind< i> ab Initio</i> Protein Structure Predictions," *Journal of structural biology,* vol. 134, no. 2, pp. 186-190, 2001.
[19] C. B. Anfinsen, "Principles that Govern the Folding of Protein Chains," 1973.
[20] J. Wang, P. Cieplak, and P. A. Kollman, "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?," *Journal of Computational Chemistry,* vol. 21, no. 12, pp. 1049-1074, 2000.
[21] E. J. Sorin, and V. S. Pande, "Exploring the helix-coil transition via all-atom equilibrium ensemble simulations," *Biophys J,* vol. 88, no. 4, pp. 2472-2493, 2005.
[22] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides," *The Journal of Physical Chemistry B,* vol. 105, no. 28, pp. 6474-6487, 2001.
[23] A. Jagielska, L. Wroblewska, and J. Skolnick, "Protein model refinement using an optimized physics-based all-atom force field," *Proceedings of the National Academy of Sciences,* vol. 105, no. 24, pp. 8268-8273, 2008.
[24] C. M. Summa, and M. Levitt, "Near-native structure refinement using in vacuo energy minimization," *Proceedings of the National Academy of Sciences,* vol. 104, no. 9, pp. 3177-3182, 2007.
[25] H. A. Scheraga, "Recent developments in the theory of protein folding: searching for the global energy minimum," *Biophysical chemistry,* vol. 59, no. 3, pp. 329-339, 1996.
[26] A. Kolinski, "Protein modeling and structure prediction with a reduced representation," *ACTA BIOCHIMICA POLONICA-ENGLISH EDITION-,* vol. 51, pp. 349-372, 2004.
[27] M. R. Lee, J. Tsai, D. Baker, and P. A. Kollman, "Molecular dynamics in the endgame of protein structure prediction," *Journal of molecular biology,* vol. 313, no. 2, pp. 417-430, 2001.
[28] H. Fan, and A. E. Mark, "Refinement of homology-based protein structures by molecular dynamics simulation techniques," *Protein Science,* vol. 13, no. 1, pp. 211-220, 2004.
[29] M. R. Lee, J. Tsai, D. Baker, and P. A. Kollman, "Molecular dynamics in the endgame of protein structure prediction1," *Journal of molecular biology,* vol. 313, no. 2, pp. 417-430, 2001.

[30] G. Chopra, N. Kalisman, and M. Levitt, "Consistent refinement of submitted models at CASP using a knowledge-based potential," *Proteins: Structure, Function, and Bioinformatics,* vol. 78, no. 12, pp. 2668-2678, 2010.
[31] J. Zhang, Y. Liang, and Y. Zhang, "Atomic-Level Protein Structure Refinement Using Fragment-Guided Molecular Dynamics Conformation Sampling," *Structure,* vol. 19, no. 12, pp. 1784-1795, 2011.
[32] D. Xu, and Y. Zhang, "Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-Step Atomic-Level Energy Minimization," *Biophys J,* vol. 101, no. 10, pp. 2525, 2011.
[33] J. Kopp, L. Bordoli, J. N. D. Battey, F. Kiefer, and T. Schwede, "Assessment of CASP7 predictions for template-based modeling targets," *Proteins: Structure, Function, and Bioinformatics,* vol. 69, no. S8, pp. 38-56, 2007.
[34] D. A. Keedy, C. J. Williams, J. J. Headd, W. B. Arendall, V. B. Chen, G. J. Kapral, R. A. Gillespie, J. N. Block, A. Zemla, and D. C. Richardson, "The other 90% of the protein: Assessment beyond the Cαs for CASP8 template-based and high-accuracy models," *Proteins: Structure, Function, and Bioinformatics,* vol. 77, no. S9, pp. 29-49, 2009.
[35] V. Mariani, F. Kiefer, T. Schmidt, J. Haas, and T. Schwede, "Assessment of template based protein structure predictions in CASP9," *Proteins: Structure, Function, and Bioinformatics*, 2011.
[36] D. Cozzetto, A. Kryshtafovych, K. Fidelis, J. Moult, B. Rost, and A. Tramontano, "Evaluation of template-based models in CASP8 with standard measures," *Proteins: Structure, Function, and Bioinformatics,* vol. 77, no. S9, pp. 18-28, 2009.
[37] R. Jauch, H. C. Yeo, P. R. Kolatkar, and N. D. Clarke, "Assessment of CASP7 structure predictions for template free targets," *Proteins: Structure, Function, and Bioinformatics,* vol. 69, no. S8, pp. 57-67, 2007.
[38] M. Ben-David, O. Noivirt-Brik, A. Paz, J. Prilusky, J. L. Sussman, and Y. Levy, "Assessment of CASP8 structure predictions for template free targets," *Proteins: Structure, Function, and Bioinformatics,* vol. 77, no. S9, pp. 50-65, 2009.
[39] C. Floudas, H. Fung, S. McAllister, M. Mönnigmann, and R. Rajgaria, "Advances in protein structure prediction and de novo protein design: A review," *Chemical Engineering Science,* vol. 61, no. 3, pp. 966-988, 2006.
[40] L. Kinch, S. Yong Shi, Q. Cong, H. Cheng, Y. Liao, and N. V. Grishin, "CASP9 assessment of free modeling target predictions," *Proteins: Structure, Function, and Bioinformatics*, 2011.
[41] J. L. MacCallum, L. Hua, M. J. Schnieders, V. S. Pande, M. P. Jacobson, and K. A. Dill, "Assessment of the protein-structure refinement category in CASP8," *Proteins: Structure, Function, and Bioinformatics,* vol. 77, no. S9, pp. 66-80, 2009.
[42] J. L. MacCallum, A. Pérez, M. J. Schnieders, L. Hua, M. P. Jacobson, and K. A. Dill, "Assessment of protein structure refinement in CASP9," *Proteins: Structure, Function, and Bioinformatics,* vol. 79, no. S10, pp. 74-90, 2011.

[43] J. Chen, and C. L. Brooks, "Can molecular dynamics simulations provide high-resolution refinement of protein structure?," *Proteins: Structure, Function, and Bioinformatics,* vol. 67, no. 4, pp. 922-930, 2007.
[44] H. Fan, and A. E. Mark, "Refinement of homology-based protein structures by molecular dynamics simulation techniques," *Protein Science,* vol. 13, no. 1, pp. 211-220, 2009.
[45] R. Ishitani, T. Terada, and K. Shimizu, "Refinement of comparative models of protein structure by using multicanonical molecular dynamics simulations," *Molecular Simulation,* vol. 34, no. 3, pp. 327-336, 2008.
[46] K. Misura, and D. Baker, "Progress and challenges in high-resolution refinement of protein structure models," *Proteins: Structure, Function, and Bioinformatics,* vol. 59, no. 1, pp. 15-29, 2005.
[47] B. D. Sellers, K. Zhu, S. Zhao, R. A. Friesner, and M. P. Jacobson, "Toward better refinement of comparative models: predicting loops in inexact environments," *Proteins: Structure, Function, and Bioinformatics,* vol. 72, no. 3, pp. 959-971, 2008.
[48] J. Zhu, H. Fan, X. Periole, B. Honig, and A. E. Mark, "Refining homology models by combining replica-exchange molecular dynamics and statistical potentials," *Proteins: Structure, Function, and Bioinformatics,* vol. 72, no. 4, pp. 1171-1188, 2008.
[49] L. Wroblewska, A. Jagielska, and J. Skolnick, "Development of a physics-based force field for the scoring and refinement of protein models," *Biophysical journal,* vol. 94, no. 8, pp. 3227-3240, 2008.
[50] A. Sali, and T. Blundell, "Comparative protein modelling by satisfaction of spatial restraints," *Protein Structure by Distance Analysis,* vol. 64, pp. C86, 1994.
[51] D. Fischer, "3D-SHOTGUN: A novel, cooperative, fold-recognition meta-predictor," *Proteins: Structure, Function, and Bioinformatics,* vol. 51, no. 3, pp. 434-441, 2003.
[52] J. Cheng, "A multi-template combination algorithm for protein comparative modeling," *BMC Structural Biology,* vol. 8, no. 1, pp. 18, 2008.
[53] S. Wu, and Y. Zhang, "LOMETS: a local meta-threading-server for protein structure prediction," *Nucleic acids research,* vol. 35, no. 10, pp. 3375-3382, 2007.
[54] Y. Zhang, "Template-based modeling and free modeling by I-TASSER in CASP7," *Proteins: Structure, Function, and Bioinformatics,* vol. 69, no. S8, pp. 108-117, 2007.
[55] K. Joo, J. Lee, S. Lee, J. H. Seo, and S. J. Lee, "High accuracy template based modeling by global optimization," *Proteins: Structure, Function, and Bioinformatics,* vol. 69, no. S8, pp. 83-89, 2007.
[56] K. M. S. Misura, D. Chivian, C. A. Rohl, D. E. Kim, and D. Baker, "Physically realistic homology models built with ROSETTA can be more accurate than their templates," *Proceedings of the National Academy of Sciences,* vol. 103, no. 14, pp. 5361-5366, 2006.

[57] Q. Wang, A. A. Canutescu, and R. L. Dunbrack, "SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling," *Nature protocols,* vol. 3, no. 12, pp. 1832-1847, 2008.

[58] G. G. Krivov, M. V. Shapovalov, and R. L. Dunbrack Jr, "Improved prediction of protein side-chain conformations with SCWRL4," *Proteins: Structure, Function, and Bioinformatics,* vol. 77, no. 4, pp. 778-795, 2009.

[59] M. Levitt, "Accurate modeling of protein conformation by automatic segment matching," *Journal of molecular biology,* vol. 226, no. 2, pp. 507-533, 1992.

[60] N. Eswar, B. Webb, M. A. Marti-Renom, M. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper, and A. Sali, "Comparative protein structure modeling using Modeller," *Curr Protoc Protein Sci,* vol. 2, no. 12, pp. 15-32, 2007.

[61] B. Qian, S. Raman, R. Das, P. Bradley, A. J. McCoy, R. J. Read, and D. Baker, "High-resolution structure prediction and the crystallographic phase problem," *Nature,* vol. 450, no. 7167, pp. 259-264, 2007.

[62] J. L. MacCallum, A. Pérez, M. J. Schnieders, L. Hua, M. P. Jacobson, and K. A. Dill, "Assessment of protein structure refinement in CASP9," *Proteins: Structure, Function, and Bioinformatics*, 2011.

[63] A. Jagielska, L. Wroblewska, and J. Skolnick, "Protein model refinement using an optimized physics-based all-atom force field," *Proceedings of the National Academy of Sciences,* vol. 105, no. 24, pp. 8268, 2008.

[64] C. M. Summa, and M. Levitt, "Near-native structure refinement using in vacuo energy minimization," *Proceedings of the National Academy of Sciences,* vol. 104, no. 9, pp. 3177, 2007.

[65] N. Kalisman, A. Levi, T. Maximova, D. Reshef, S. Zafriri-Lynn, Y. Gleyzer, and C. Keasar, "MESHI: a new library of Java classes for molecular modeling," *Bioinformatics,* vol. 21, no. 20, pp. 3931-3932, 2005.

[66] G. D. Rose, P. J. Fleming, J. R. Banavar, and A. Maritan, "A backbone-based theory of protein folding," *Proceedings of the National Academy of Sciences,* vol. 103, no. 45, pp. 16623-16633, 2006.

[67] K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, "Principles of protein folding--a perspective from simple exact models," *Protein science: a publication of the Protein Society,* vol. 4, no. 4, pp. 561, 1995.

[68] S. Wallin, and E. I. Shakhnovich, "Understanding ensemble protein folding at atomic detail," *Journal of Physics: Condensed Matter,* vol. 20, pp. 283101, 2008.

[69] N. Engler, A. Ostermann, N. Niimura, and F. G. Parak, "Hydrogen atoms in proteins: Positions and dynamics," *Proceedings of the National Academy of Sciences,* vol. 100, no. 18, pp. 10243, 2003.

[70] I. K. McDonald, and J. M. Thornton, "Satisfying hydrogen bonding potential in proteins," *Journal of molecular biology,* vol. 238, no. 5, pp. 777-793, 1994.

[71] J. Chen, and C. L. Brooks III, "Can molecular dynamics simulations provide high-resolution refinement of protein structure?," *Proteins: Structure, Function, and Bioinformatics,* vol. 67, no. 4, pp. 922-930, 2007.

[72] G. Vriend, "WHAT IF: a molecular modeling and drug design program," *Journal of molecular graphics,* vol. 8, no. 1, pp. 52, 1990.

[73] A. T. Brünger, and M. Karplus, "Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison," *Proteins: Structure, Function, and Bioinformatics,* vol. 4, no. 2, pp. 148-156, 1988.
[74] J. M. Word, S. C. Lovell, J. S. Richardson, and D. C. Richardson, "Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation1," *Journal of molecular biology,* vol. 285, no. 4, pp. 1735-1747, 1999.
[75] Y. Li, A. Roy, and Y. Zhang, "HAAD: a quick algorithm for accurate prediction of hydrogen atoms in protein structures," *PloS one,* vol. 4, no. 8, pp. e6701, 2009.
[76] M. Levitt, M. Hirshberg, R. Sharon, and V. Daggett, "Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution," *Computer Physics Communications,* vol. 91, no. 1, pp. 215-231, 1995.
[77] W. Kabsch, and C. Sander, "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers,* vol. 22, no. 12, pp. 2577-2637, 1983.
[78] E. A. D. Amir, N. Kalisman, and C. Keasar, "Differentiable, multi-dimensional, knowledge-based energy terms for torsion angle probabilities and propensities," *Proteins: Structure, Function, and Bioinformatics,* vol. 72, no. 1, pp. 62-73, 2008.
[79] M. Friedel, A. Baumketner, and J. E. Shea, "Effects of surface tethering on protein folding mechanisms," *Proceedings of the National Academy of Sciences,* vol. 103, no. 22, pp. 8396-8401, 2006.
[80] D. C. Liu, and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical programming,* vol. 45, no. 1, pp. 503-528, 1989.
[81] J. Moult, T. Hubbard, K. Fidelis, and J. T. Pedersen, "Critical assessment of methods of protein structure prediction (CASP): round III," *Proteins: Structure, Function, and Bioinformatics,* vol. 37, no. S3, pp. 2-6, 1999.
[82] Z. Wang, J. Eickholt, and J. Cheng, "MULTICOM: a multi-level combination approach to protein structure prediction and its assessments in CASP8," *Bioinformatics,* vol. 26, no. 7, pp. 882-888, 2010.
[83] A. Zemla, "LGA: a method for finding 3D similarities in protein structures," *Nucleic acids research,* vol. 31, no. 13, pp. 3370-3374, 2003.
[84] Y. Zhang, and J. Skolnick, "Scoring function for automated assessment of protein structure template quality," *Proteins: Structure, Function, and Bioinformatics,* vol. 57, no. 4, pp. 702-710, 2004.
[85] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography,* vol. 32, no. 5, pp. 922-923, 1976.
[86] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson, "MolProbity: all-atom structure validation for macromolecular crystallography," *Acta Crystallographica Section D: Biological Crystallography,* vol. 66, no. 1, pp. 12-21, 2009.
[87] D. A. Keedy, C. J. Williams, J. J. Headd, W. B. Arendall III, V. B. Chen, G. J. Kapral, R. A. Gillespie, J. N. Block, A. Zemla, and D. C. Richardson, "The other 90% of the protein: Assessment beyond the Cαs for CASP8 template-based and

high-accuracy models," *Proteins: Structure, Function, and Bioinformatics,* vol. 77, no. S9, pp. 29-49, 2009.

[88] M. Levitt, and M. Gerstein, "A unified statistical framework for sequence comparison and structure comparison," *Proceedings of the National Academy of Sciences,* vol. 95, no. 11, pp. 5913, 1998.

[89] J. Xu, and Y. Zhang, "How significant is a protein structure similarity with TM-score= 0.5?," *Bioinformatics,* vol. 26, no. 7, pp. 889-895, 2010.

[90] S. C. Lovell, J. M. Word, J. S. Richardson, and D. C. Richardson, "The penultimate rotamer library," *Proteins: Structure, Function, and Bioinformatics,* vol. 40, no. 3, pp. 389-408, 2000.

[91] W. B. Arendall III, W. Tempel, J. S. Richardson, W. Zhou, S. Wang, I. W. Davis, Z.-J. Liu, J. P. Rose, W. M. Carson, and M. Luo, "A test of enhancing model accuracy in high-throughput crystallography," *Journal of structural and functional genomics,* vol. 6, no. 1, pp. 1-11, 2005.

[92] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, and J. S. Richardson, "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids," *Nucleic acids research,* vol. 35, no. suppl 2, pp. W375-W383, 2007.

[93] A. Kryshtafovych, Č. Venclovas, K. Fidelis, and J. Moult, "Progress over the first decade of CASP experiments," *Proteins: Structure, Function, and Bioinformatics,* vol. 61, no. S7, pp. 225-236, 2005.

[94] O. Keskin, and I. Bahar, "Packing of sidechains in low-resolution models for proteins," *Folding and Design,* vol. 3, no. 6, pp. 469-479, 1998.

[95] H. Wieman, T. Kristin, E. Anderssen, and F. Drablos, "Homology-based modelling of targets for rational drug design," *Mini reviews in medicinal chemistry,* vol. 4, no. 7, pp. 793-804, 2004.

[96] R. Méndez, R. Leplae, M. F. Lensink, and S. J. Wodak, "Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures," *Proteins: Structure, Function, and Bioinformatics,* vol. 60, no. 2, pp. 150-169, 2005.

[97] A. K. Arakaki, Y. Zhang, and J. Skolnick, "Large-scale assessment of the utility of low-resolution protein structures for biochemical function assignment," *Bioinformatics,* vol. 20, no. 7, pp. 1087-1096, 2004.

[98] D. Bhattacharya, and J. Cheng, "3Drefine: Consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization," *Proteins: Structure, Function, and Bioinformatics*, 2012.

[99] K. Olechnovič, E. Kulberkytė, and Č. Venclovas, "CAD-score: A new contact area difference-based function for evaluation of protein structural models," *Proteins: Structure, Function, and Bioinformatics*, 2012.

[100] R. A. Abagyan, and M. M. Totrov, "Contact area difference (CAD): a robust measure to evaluate accuracy of protein models," *Journal of molecular biology,* vol. 268, no. 3, pp. 678-685, 1997.

[101] D. Baker, and A. Sali, "Protein structure prediction and structural genomics," *Science's STKE,* vol. 294, no. 5540, pp. 93, 2001.

[102] N. Siew, A. Elofsson, L. Rychlewski, and D. Fischer, "MaxSub: an automated measure for the assessment of protein structure prediction quality," *Bioinformatics,* vol. 16, no. 9, pp. 776-785, 2000.
[103] M. Fasnacht, J. Zhu, and B. Honig, "Local quality assessment in homology models using statistical potentials and support vector machines," *Protein Science,* vol. 16, no. 8, pp. 1557-1568, 2007.
[104] P. Larsson, M. J. Skwark, B. Wallner, and A. Elofsson, "Assessment of global and local model quality in CASP8 using Pcons and ProQ," *Proteins: Structure, Function, and Bioinformatics,* vol. 77, no. S9, pp. 167-172, 2009.
[105] B. Wallner, and A. Elofsson, "Identification of correct regions in protein models using structural, alignment, and consensus information," *Protein Science,* vol. 15, no. 4, pp. 900-913, 2006.
[106] A. Fiser, and A. Šali, "Modeller: generation and refinement of homology-based protein structure models," *Methods in enzymology,* vol. 374, pp. 461-491, 2003.
[107] J. Zhang, and Y. Zhang, "A novel side-chain orientation dependent potential derived from random-walk reference state for protein fold selection and structure prediction," *PloS one,* vol. 5, no. 10, pp. e15386, 2010.
[108] Y. Yang, and Y. Zhou, "Specific interactions for ab initio folding of protein terminal regions with secondary structures," *Proteins: Structure, Function, and Bioinformatics,* vol. 72, no. 2, pp. 793-803, 2008.
[109] Y. Yang, and Y. Zhou, "Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions," *Protein Science,* vol. 17, no. 7, pp. 1212-1219, 2008.
[110] M. Shen, and A. Sali, "Statistical potential for assessment and prediction of protein structures," *Protein Science,* vol. 15, no. 11, pp. 2507-2524, 2006.
[111] S. C. E. Tosatto, "The victor/FRST function for model quality estimation," *Journal of Computational Biology,* vol. 12, no. 10, pp. 1316-1327, 2005.
[112] R. Samudrala, and J. Moult, "An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction1," *Journal of molecular biology,* vol. 275, no. 5, pp. 895-916, 1998.
[113] S. C. E. Tosatto, and R. Battistutta, "TAP score: torsion angle propensity normalization applied to local protein structure evaluation," *BMC bioinformatics,* vol. 8, no. 1, pp. 155, 2007.
[114] Z. Wang, A. N. Tegge, and J. Cheng, "Evaluating the absolute quality of a single protein model using structural features and support vector machines," *Proteins: Structure, Function, and Bioinformatics,* vol. 75, no. 3, pp. 638-647, 2009.
[115] V. Pihur, and S. Datta, "RankAggreg, an R package for weighted rank aggregation," *BMC bioinformatics,* vol. 10, no. 1, pp. 62, 2009.
[116] R. Rubinstein, "The cross-entropy method for combinatorial and continuous optimization," *Methodology and computing in applied probability,* vol. 1, no. 2, pp. 127-190, 1999.
[117] P. T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Annals of operations research,* vol. 134, no. 1, pp. 19-67, 2005.

[118] L. M. K. Adler, "A modification of Kendall's tau for the case of arbitrary ties in both rankings," *Journal of the American Statistical Association,* vol. 52, no. 277, pp. 33-35, 1957.
[119] S. Wu, J. Skolnick, and Y. Zhang, "Ab initio modeling of small proteins by iterative TASSER simulations," *BMC biology,* vol. 5, no. 1, pp. 17, 2007.
[120] Y. Zhang, and J. Skolnick, "SPICKER: A clustering approach to identify near-native protein folds," *Journal of Computational Chemistry,* vol. 25, no. 6, pp. 865-871, 2004.
[121] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of chemical theory and computation,* vol. 4, no. 3, pp. 435-447, 2008.
[122] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," *Journal of the American Chemical Society,* vol. 118, no. 45, pp. 11225-11236, 1996.
[123] M. Larkin, G. Blackshields, N. Brown, R. Chenna, P. McGettigan, H. McWilliam, F. Valentin, I. Wallace, A. Wilm, and R. Lopez, "Clustal W and Clustal X version 2.0," *Bioinformatics,* vol. 23, no. 21, pp. 2947-2948, 2007.

# VITA

Debswapna Bhattacharya was born in the city of Kolkata, India. After finishing his undergraduate studies in 2005 from West Bengal University of Technology, India, majoring in Computer Science and Engineering, he opted for a professional career in the industry and worked for five years in the area of data mining, data warehousing and big-data applications mostly in Fortune 500 companies like Accenture, Cognizant etc., providing solutions to a vast range of industry verticals like retail, healthcare, manufacturing, finance and pharmaceuticals. In Fall 2011, Debswapna joined the Department of Computer Science in University of Missouri – Columbia to pursue his M.S. and Ph.D. studies under the advise of Professor Jianlin Cheng. His research is focused on applying machine learning and data mining techniques to analyze big biomedical data and address fundamental problems in biomedical sciences and he enjoys taking approaches that combine computational optimization and statistical methods with bioinformatics and systems biology.