Published in final edited form as:

Methods Mol Biol. 2012 ; 815: 3–13. doi:10.1007/978-1-61779-424-7_1.

Prediction of protein tertiary structures using MUFOLD

Jingfen Zhang^{1,2}, Zhiquan He^{1,2}, Qingguo Wang¹, Bogdan Barz³, Ioan Kosztin³, Yi Shang¹, and Dong Xu^{1,2}

¹Department of Computer Science, University of Missouri, Columbia, Missouri

²Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, Missouri

³Department of Physics and Astronomy, University of Missouri, Columbia, Missouri

Abstract

There have been steady improvements in protein structure prediction during the past two decades. However, current methods are still far from consistently predicting structural models accurately with computing power accessible to common users. To address this challenge, we developed MUFOLD, a hybrid method of using whole and partial template information along with new computational techniques for protein tertiary structure prediction. MUFOLD covers both template-based and *ab initio* predictions using the same framework and aims to achieve high accuracy and fast computing. Two major novel contributions of MUFOLD are graph-based model generation and molecular dynamics ranking (MDR). By formulating prediction as a graph realization problem, we apply an efficient optimization approach of Multidimensional Scaling (MDS) to speed up the prediction dramatically. In addition, under this framework, we enhance the predictions consistently by iteratively using the information from generated models. MDR, in contrast to widely used static scoring functions, exploits dynamics properties of structures to evaluate their qualities, which can often identify best structures from a pool more effectively.

Keywords

Protein structure prediction; Multidimensional scaling; Molecular dynamics simulation

1. Introduction

Protein tertiary structure often provides a basis for understanding its function. Experimental approaches for protein structure determination, such as X-ray crystallography (1) and Nuclear Magnetic Resonance (NMR) techniques (2), are typically expensive and time-consuming. The increase of the structures in Protein Data Bank (PDB) (3) cannot keep up with the increase of proteins characterized in high-throughput genome sequencing (4). Compared to experimental approaches, computational methods, i.e., to predict the native structure of a protein from its amino acid sequence, are much cheaper and faster. As significant progress has been made over the past two decades, computational methods are becoming more and more important for studying protein structures in recent years.

The foundation to predict the protein structure by computational methods relies on two sets of principles: the laws of physics and the theory of evolution. The protein folding theory based on the laws of physics states that at physiological conditions (temperature, ion concentration, etc.), a protein folds into its native structure with a unique, stable and kinetically accessible minimum of free energy (5). The theory of evolution gives us the other guidance for structure prediction: 1) proteins with similar sequences usually have similar structures and 2) protein structures are more conserved than their sequences (6). When the structure of one protein in a family of proteins with similar sequences/structures has been determined by experiment, the other members of the family can be modeled based on their alignments to the known structure.

According to the above foundations, computational prediction methods can be classified into three categories: 1) *ab initio* prediction (7–10), 2) comparative modeling (CM) (11–13), and 3) threading (14–17). *Ab initio* methods assume that native structure corresponds to the global free energy minimum accessible during the lifespan of the protein and attempt to find this minimum by an exploration of many conceivable protein conformations. The prediction results are unreliable due to: 1) the huge conformational search space, and 2) the limitations of the currently used scoring functions. Both CM and threading are template-based methods. CM, using sequence comparison, is a successful category of prediction methods. With the increasing accumulation of experimentally determined protein structures and the advances in remote homology identification, CM has made continuing progress. However, when the sequence identity drops below 30%, the accuracy of CM sharply decreases because of substantial alignment errors. Threading is based on sequence-structure comparison and measures the fitness of the target sequence into templates. A special category of threading called mini-threading, obtains matches between a query sequence and short structure fragments in PDB to build local structures, which are then assembled into final models that require a significantly smaller computational search space than *ab initio* methods. Thus, mini-threading has a better chance to achieve high prediction accuracy than CM in cases when no evolutionarily relationship is available between the target and template sequences.

Although significant progress has been made, existing computational methods are still far from consistently providing accurate structural models with reasonable computing time. Currently, the most popular optimization methods used in structure prediction such as genetic algorithms and Monte Carlo simulations are time-consuming so as to generate structural models often far from the global optimal solution of a scoring function. In addition, widely used scoring functions are generally not accurate enough to identify the best structure from the generated structure pool. Hence, although a number of prediction servers such as Modeller (12), HHpred (13), I-TASSA (18) and Rosetta (19), have been developed, protein structure prediction has not been widely applied in molecular biology studies other than homology modeling with structural templates of high-sequence identity, due to low prediction accuracies and long computing time.

To address the above issues, we proposed a hybrid method, MUFOLD, by using whole and partial template information to cover both template-based and *ab initio* predictions using the same framework. The framework generates structural models very fast so that it can assess and improve the model quality more directly than sequence alignment only. Two major

novel contributions of MUFOLD are fast graph-based model generation and molecular dynamics ranking (MDR).

On one hand, instead of using the Monte Carlo method to sample the conformation space, we have tried to find suitable templates and fragment structures in PDB to estimate the spatial constraints between residues in target sequence, which decreases the search space. At the same time, we bypassed the energy functions and formulated the structure prediction problem as a graph realization problem. Then we have applied an efficient optimization approach of MDS to speed up the prediction dramatically. In addition, under this graph-based framework, we can improve the distance constraints by iteratively using the information from the models and thus enhance the predictions consistently.

On the other hand, in contrast to widely used static scoring functions, we have proposed a ranking method, MDR, to exploits dynamics properties of structures to evaluate their qualities, which can often identify the best structures from a pool more effectively. This is a rare success in applications of molecular dynamics simulation for general protein structure predictions.

2. Materials

2.1. Alignment tools

In MUFOLD we make use of sequence-profile alignment tools, e.g., PSI-BLAST (20), profile-profile alignment tool, e.g., HHSearch (21) and an in-house threading approach, PROSPECT (22), to search possible templates against PDB for target sequences.

2.2. Scoring functions

Currently, structure quality assessment and model selection generally use the scoring functions in two categories (23): physics-based energy functions and knowledge-based statistical potentials. The knowledge-based statistical potentials are typically fast to calculate, easy to construct, and hence are most widely used in structure quality assessment. We investigated some state-of-the-art scoring functions, and finally chose OPUS (24), Model Evaluator (25), and Dfire energy (26) as the scoring functions to evaluate the quality of models.

2.3. Multi-Dimensional Scaling (MDS)

MDS method is efficient for solving the graph realization problem. It starts with one or more distance matrices derived from points in a multidimensional space and finds a placement of the points in a low-dimensional space. In MUFOLD, we estimate the distances between C α (or backbone) atoms for each pair of amino acids in the target sequence as distance matrix and then calculate the coordinates of the atom for each amino acid. We generate models using different techniques of MDS: classical metric MDS (CMDS) (27), weighted MDS (WMDS) (28), and split-and-combine MDS (SC-MDS) (29). In our study, we mainly use CMDS, which is the simplest MDS algorithm. CMDS minimizes the sum of least squared errors between the estimated distances and the actual distances in the output model for all pairs of points.

3. Methods

3.1. The overview of MUFOLD

MUFOLD takes whole and partial template information for both template-based and *ab initio* predictions using the same framework towards achieving improved accuracies and fast computing in automated predictions. The overview of MUFOLD is presented in Fig. 1, which includes three main parts: 1) template selection and alignment, i.e., recognizing potentially useful templates/fragments in PDB for the target sequence and building alignments; 2) coarse-grain model generations and evaluations, including fast model generations using MDS techniques at the C α or backbone level, evaluations of models through static scoring functions, and iterative improvement of selected models by integrating spatial constraints from sequence alignments and selected models; and 3) full-atom model evaluation through MD simulations.

3.2. Template selection and alignment

The first step of MUFOLD is to find suitable templates and alignments. Here, the “template” is a general concept including the global homologous templates, non-evolutionarily related (analogous) templates, and the locally compatible protein fragments from PDB. MUFOLD adaptively applies different strategies for various targets. In general, target sequences are classified into three categories: Easy, Medium, and Hard.

“Easy” targets have significant hits by applying sequence-profile alignment in PSI-BLAST (20) against the PDB, i.e., there is at least one alignment hit that can cover more than 70% of the target sequence and with E-value $1e-3$ or less. As homologous templates with high confidence alignments can be easily found for this case, it is intuitive that the sequence alignment can be used to obtain high-quality distance constraints directly. “Medium” targets have remote homologies obtained by using profile-profile alignment in HHSearch (21), i.e., there is at least one alignment hit with E-value less than $1e-2$ (excluding “easy” targets defined above). These targets probably have the correctly identified fold information, but the alignments may be incorrect. Therefore, we try to obtain various alignments by applying different tools and parameters for the correct fold. Coupled with the optimization of MDS, we sample distance constraints and improve the constraints iteratively. “Hard” targets have analog structural templates in PDB that cannot be assigned even by profile-profile alignment. We use an in-house threading approach, PROSPECT (22), to search for possible templates. Although the top one hit may not represent the correct fold, the compatible protein fragments of top n (20–100) folds usually include the correct fold.

3.3. Coarse-grain model generation

3.3.1 Graph-based model generation formulation—We formulate the structure prediction problem as a graph realization problem and then apply MDS technique to solve it. The basic idea is to estimate the distances between C α atoms for each pair of residues in the target sequence and then calculate the corresponding C α coordinates by applying MDS. Assume there are n points (each representing the C α atom of a residue) $X_k \in R^3$, $k = 1, \dots, n$ in a 3-D space. If we know the exact distances between some pairs of points, e.g., d_{ij} between residue i at X_i and residue j at X_j , then the graph realization problem is to determine

the coordinates of the points from the partial distance constraints such that the distance between each pair of points matches the given distance constraint, $\|X_i - X_j\| = d_{ij}$ for all d_{ij} . If the distance constraints are inaccurate, usually there is no exact or unique solution to the over-determined system of equations. Instead, the problem is formulated as an optimization problem that minimizes the sum of squared errors as:

$$\min_{X_1 \dots X_n \in R^3} \sum_{i,j=1, \dots, n} (\|X_i - X_j\| - d_{ij})^2 \quad \text{Equation (1)}$$

The optimization problem of Equation (1) is generally non-convex with many local minima and MDS is very suitable for this optimization problem.

3.3.2 Spatial distance constraints—Since predicted distance constraints are often noisy, our strategy is to keep refining the initial models by sampling and improving the distance constraints (or contact maps) iteratively. The initial contact maps of a target protein are retrieved from alignments between the target sequence and various template proteins in PDB obtained in the above step of “template selection and alignment”.

For a given alignment between the target and a template, we first estimate the pair-wise distance of the aligned residues in target by the distance of the corresponding residues in the template. Although we select multiple long templates and short fragments, there may still be residues in the target that are aligned to gaps or two residues that are not covered by any single hit simultaneously so that related pair-wise distances cannot be derived directly. For these missing distances, we estimate them by the shortest-path distance. We know that the adjacent C α atom distance is about 3.8 Å, which means that any two C α atoms can be connected at least through adjacent C α atoms. There may be many different paths to connect two C α atoms, we use the shortest path distance to estimate the unknown pair-wise distance. Although the shortest path often over-estimates the distance, it provides an initial complete contact map for calculating a model by MDS.

It should be mentioned that MDS generates two mirror models for any given contact map. Technically, we superimpose the model configuration to the template, and calculate the reflection factor of the superimposition. If the reflection factor equals to 1, it indicates that the configuration is correct; otherwise it is the incorrect mirror.

3.4 Coarse-grain model evaluation

Coarse-grain model generation using MDS leads to a large number of candidate structures. In MUFOLD, we apply static scorings to evaluate and select better models for the next iteration of model improvement. Specifically, the method consists of filtering and representative finding. At first, we calculate the scoring functions, such as OPUS, Model Evaluator, and Dfire energy for each model. These scoring functions are normalized to z-score and summed to filter out those models with lower sum value. Next, the remaining structures are grouped into clusters based on pair-wise similarity measured by RMSD. For each cluster, we find a representative model whose average RMSD to all the other models in

the cluster is minimal. These representative models can be reported as final models or as the input models for the next iteration.

3.5 Iterative improving process

Although using multiple templates and fragments can generate models that are closer to the native structure than any template alone, inconsistent constraints from different alignments and distances estimated by the shortest path method may compromise the quality of the models. Our strategy is to refine and improve the constraints iteratively by combining the original constraints derived from the alignments ($D_{\text{alignment}}$) and the measured distances from the generated models (D_{model}) as: $D_{\text{refine}} = \lambda * D_{\text{alignment}} + (1 - \lambda) * D_{\text{model}}$, $0 \leq \lambda \leq 1$. There are different ways to set the value of λ . For example, a simple way is to set $\lambda = 0.5$ if $D_{\text{alignment}}$ is available, otherwise $\lambda = 0$. Another way is to set λ according to the confidence level of $D_{\text{alignment}}$. By performing this iterative generation, the quality of models often gets better and better, while many deficiencies in the models are fixed over iterations.

Figure 2 shows an example of iteratively improved coarse-grain model generation, where we show the original and improved contact maps and the corresponding models in (a)-(c), (d)-(f), respectively. In the image of contact map, we use colors to illustrate the distances between pairwise residues, where the lighter the color is, the larger the distance is. We can observe the color changes within the red rectangle regions in Fig. 2 (a) and (d), which means the modification of the distance constraints. From the data showed in Fig. 2 (b) and (e), we can see the significant improvement of the models, for example, all of the quality score such as RMSD, GDT_TS, GDT_HA(30) and TMscore (31) of the model against the native have been improved.

3.6 Full-atom model evaluation: MD-Ranking (MDR)

The coarse-grain model generation described above provides various structures with significantly different conformations. How to identify the one with the smallest RMSD compared to the unknown native structure is a highly challenging problem. Existing methods generally use static scoring functions (measurements from static conformations) to rank models. However, the dynamics properties of a model may reveal its structural quality better than static information. Near native models are always more stable than poor-quality models during simulated heating, i.e., the latter unfold at lower temperatures than the former. Thus, the quantitative assessment of relative stabilities of structural models against gradual heating provides an alternative way of ranking the structures' quality. Here, we propose a novel MD-Ranking (MDR) method based on full-atom MD simulations (32) to evaluate and rank protein models according to their stabilities against external perturbations, e.g., change in temperature or externally applied forces. The basic idea is to build all-atom models from the coarse-grain models, optimize these models by energy minimization, gradually heat them through MD simulations, and then rank the models based on their structural changes during heating.

More specifically, first, an all-atom model is built for each of the top selected structures by the above coarse-grain model process. The coordinates of the missing backbone and side-chain heavy atoms are predicted by using the program Pulchra (33), and the hydrogen atoms

are added by using psfgen, which is part of the VMD package (34). Next, the obtained structures are optimized by removing the bad contacts through energy minimization. Finally, the stability of a structure is tested by monitoring the change of its C α RMSD (cRMSD) with respect to its initial structure during the MD simulation of a scheduled heating at a rate of 1K/ps. The MD simulations are carried out in vacuum by coupling the system to a Langevin heat bath whose temperature can be varied (i.e., the dynamics of protein atoms is described by a Langevin equation). All energy minimization and MD simulations were performed by employing the CHARMM27 force field and the parallel NAMD2.6 MD simulation program (35).

Figure 3 shows three typical examples of MDR, i.e., plots of the changes in cRMSD during the heating MD simulations. For the first case (Fig. 3a), the data set contains a good model with RMSD < 3 Å to the native and many poor models. In this case, MDR can easily differentiate the best one from the others. When the best structure in the set has RMSD > 3 Å, the top ranked model of MDR is within 0.5 Å from the best one in most cases (Fig. 3b). In a few cases, however, when the quality of the models are similar and not good enough, MDR method yielded only mediocre results, as shown in Fig. 3c, where the curves of different cRMSD changes mostly overlap with lack of discerning power. In summary, the performance of MDR varies for different cases while it is most efficient when the pool of models contains high-quality models (RMSD < 3 Å) besides poor ones.

4. Notes

As a completely new framework for protein structure prediction, there are various limitations to address and new functionalities to implement for MUFOLD. MUFOLD currently can only handle protein monomers but not protein oligomers or complexes. We are improving the system in many aspects. For example, we are using multiple sequence alignment information to improve the distance constraints. Furthermore, the lack of solvent in the MD simulations may lead to errors of ranking, especially for structures that show comparable change in cRMSD during heating. Therefore, the MDR ranking method can be improved by considering a longer heating interval, using the GDT-TS or TM score instead of cRMSD, and including implicit solvent in the simulation, although adding implicit solvent may not be feasible in large-scale protein structure prediction due to long computational time. Like other tools, it is important to combine predicted structural models and wet-lab experiments to take advantage of the power of protein structure prediction.

Acknowledgments

This work has been supported by National Institutes of Health Grant R21/R33-GM078601. Major computing resource was provided by the University of Missouri Bioinformatics Consortium. We like to thank Jianlin Cheng, Yang Zhang, and Joel L. Sussman for helpful discussions.

References

1. Browne WJ, North AC, Phillips DC, Brew K, Vanaman TC, Hill RL. A possible three-dimensional structure of bovine alpha-lactalbumin based on that of hen's egg-white lysozyme. *J. Mol. Biol.* 1969; 42:65–86. [PubMed: 5817651]
2. Wuthrich K. The way to NMR structures of proteins. *Nature Structural Biology.* 2001; 8:923–925.

3. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res.* 2000; 28:235–242. [PubMed: 10592235]
4. The UniProt Consortium. The Universal Protein Resource (UniProt). *Nucleic Acids Res.* 2008; 36:D190–D195. [PubMed: 18045787]
5. Anfinsen C. The formation and stabilization of protein structure. *Biochem J.* 128(4):737–749. [PubMed: 4565129]
6. Browne WJ, North AC, Phillips DC, Brew K, Vanaman TC, Hill RL. A possible three-dimensional structure of bovine alpha-lactalbumin based on that of hen's egg-white lysozyme. *J. Mol. Biol.* 1969; 42:65–86. [PubMed: 5817651]
7. HA. Monte carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci.* 1987; 84:6611–6615. [PubMed: 3477791]
8. Liwo A, Lee J, Ripoll DR, Pillardy J, Scheraga HA. Protein structure prediction by global optimization of a potential energy function. *Proc. Natl. Acad. Sci.* 1999; 96:5482–5485. [PubMed: 10318909]
9. Simons KT, Strauss C, Baker D. Prospects for ab initio protein structural genomics. *J. Mol. Biol.* 2001; 306:1191–1199. [PubMed: 11237627]
10. Zhang Y, Kolinski A, Skolnick J. TOUCHSTONE II: A New Approach to Ab Initio Protein Structure Prediction. *Biophys J.* 2003; 85:1145–1164. [PubMed: 12885659]
11. Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science.* 1991; 253:164–170. [PubMed: 1853201]
12. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial constraints. *J Mol Biol.* 1993; 234:779–815. [PubMed: 8254673]
13. Soding J, Biegert A, Lupas A. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research.* 2005; 33:W244–W248. [PubMed: 15980461]
14. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* 1997; 268:209–225. [PubMed: 9149153]
15. Xu Y, Xu D. Protein threading using PROSPECT: Design and evaluation. *Proteins: Struct Funct Bioinformatics.* 2000; 40:343–354.
16. Inbar Y, Benyamini H, Nussinov R, Wolfson HJ. Protein structure prediction via combinatorial assembly of sub-structural units. *Bioinformatics.* 2003; 19:158–168.
17. Skolnick J, Kihara D, Zhang Y. Development and large scale benchmark testing of the PROSPECTOR_3 threading algorithm. *Proteins: Struct Funct Bioinformatics.* 2004; 56:502–518.
18. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics.* 2008; 9:40. [PubMed: 18215316]
19. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* 2004; 32(2):526–531.
20. Madden T, Schäffer A, Zhang J, Zhang Z, Miller W, Lipman D. Gapped BLAST and PSIBLAST: a new generation of protein database search programs. *Nucleic Acids Research.* 1997; 25(17): 3389–3402. [PubMed: 9254694]
21. Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics.* 2005; 21:951–960. [PubMed: 15531603]
22. Xu Y, Xu D. Protein threading using PROSPECT: Design and evaluation. *Proteins: Struct Funct Bioinformatics.* 2000; 40:343–354.
23. Xu, Y.; Xu, D.; Liang, J. *Computational Methods for Protein Structure Prediction and Modeling*, I, II. Springer-Verlag; 2006.
24. Wu Y, Lu M, Chen M, Li J, Ma J. OPUS-Ca: A knowledge-based potential function requiring only Ca positions. *Protein Science.* 2007; 16:1449–1463. [PubMed: 17586777]
25. Wang Z, Tegge A, Cheng J. Evaluating the absolute quality of a single protein model using structural features and support vector machines. *Proteins: Struct Funct Bioinformatics.* 2009; 75:638–647.

26. Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Science*. 2002; 11:2714–2726. [PubMed: 12381853]
27. Borg, I.; Groenen, P. *Modern Multidimensional Scaling – theory and applications*. New York: Springer-Verlag; 1997.
28. Torgerson WS. Multidimensional scaling of similarity. *Psychometrika*. 1965; 30:379–393. [PubMed: 5217606]
29. Tzeng J, Lu H, Li W. Multidimensional scaling for large genomic data sets. *BMC Bioinformatics*. 2008; 9:179. [PubMed: 18394154]
30. Zemla A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Research*. 2003; 31:3370–3374. [PubMed: 12824330]
31. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins*. 2004; 57:702–710. [PubMed: 15476259]
32. Phillips J, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005; 26:1781–1802. [PubMed: 16222654]
33. Feig M, Rotkiewicz P, Kolinski A, Skolnick J, Brooks CL 3rd. Accurate reconstruction of allatom protein representations from side-chain-based low-resolution models. *Proteins: Struct Funct Bioinformatics*. 2000; 41(1):86–97.
34. Humphrey W, Dalke A, Shulten K. VMD - Visual Molecular Dynamics. *J. Molec. Graphics*. 1996; 14:33–38.
35. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *J. Comput. Chem*. 2005; 26(16):1781–1802. [PubMed: 16222654]

Fig. 1.
Flowchart of the MUFOLD structure prediction method.

(a) Original contact map.

(b) Model against the native.
 Rmsd: 3.665 GDT_TS: 0.687
 GDT_HA: 0.485 TMscore: 0.737

(c) Trace of Cα atoms against the native.

(d) Improved contact map.

(e) New model against the native. Rmsd: 3.211 GDT_TS: 0.784
 GDT_HA: 0.595 TMscore: 0.826

(f) Trace of Cα atoms against the native for the new model.

Fig. 2.
 An example of iterative coarse-grain model generation.

Fig. 3.

Change in RMSD during the heating in MD simulations for three proteins. The colored curves correspond to the five different models with various RMSD values. The bold faced models have been top ranked by MDR. a) **black** – 2.7 Å, red – 3.9 Å, blue – 12.3 Å, orange – 12.6 Å, magenta – 12.9 Å; b) black – 3.1 Å, red – 3.2 Å, **blue** – 3.3 Å, orange – 9.9 Å, magenta – 9.9 Å; c) black – 3.3 Å, red – 4.5 Å, **blue** – 5.4 Å, orange – 6.3 Å, magenta – 6.7 Å