

# ResQ: An Approach to Unified Estimation of *B*-Factor and Residue-Specific Error in Protein Structure Prediction

Jianyi Yang<sup>1,2</sup>, Yan Wang<sup>2,3</sup> and Yang Zhang<sup>2,4</sup>

**1** - School of Mathematical Sciences, Nankai University, Tianjin 300071, China

**2** - Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA

**3** - School of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

**4** - Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109, USA

**Correspondence to Yang Zhang:** Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA. [zhng@umich.edu](mailto:zhng@umich.edu)

<http://dx.doi.org/10.1016/j.jmb.2015.09.024>

Edited by M. Sternberg

## Abstract

Computer-based structure prediction becomes a major tool to provide large-scale structure models for annotating biological function of proteins. Information of residue-level accuracy and thermal mobility (or *B*-factor), which is critical to decide how biologists utilize the predicted models, is however missed in most structure prediction pipelines. We developed ResQ for unified residue-level model quality and *B*-factor estimations by combining local structure assembly variations with sequence-based and structure-based profiling. ResQ was tested on 635 non-redundant proteins with structure models generated by I-TASSER, where the average difference between estimated and observed distance errors is 1.4 Å for the confidently modeled proteins. ResQ was further tested on structure decoys from CASP9-11 experiments, where the error of local structure quality prediction is consistently lower than or comparable to other state-of-the-art predictors. Finally, ResQ *B*-factor profile was used to assist molecular replacement, which resulted in successful solutions on several proteins that could not be solved from constant *B*-factor settings.

© 2015 Elsevier Ltd. All rights reserved.

## Introduction

With the rapidly increasing gap between the number of protein sequences and the number of experimentally characterized structures, computer-based structure prediction becomes a major means for the molecular and cellular biologists to obtain three-dimensional structure models of proteins for interpreting the biological function or designing new biochemical experiments. Although progress has been constantly witnessed on structure predictions in the community-wide CASP experiments [1], a reliable estimation of the quality of predicted structure models, in particular, the residue-level local accuracy that is critical to the structure-based functional analyses (such as active-site recognition, ligand docking and drug screening), is often missed in most of the state-of-the-art structure prediction pipelines [2].

Another highly relevant but often-missed local feature in structure prediction is the inherent thermal

mobility of protein atoms. At the absolute zero of temperature, protein atoms are assumed to stay at the equilibrium position of the lowest energy. However, as the temperature increases, the ambient thermal energy causes the atoms to oscillate around the equilibrium positions, the extent of which often varies depending on the three-dimensional structure and the interaction with ligand and solvent atoms. The atomic motion can be experimentally measured in the X-ray crystallography as *B*-factor (or temperature factor), which was introduced as an amendment factor of the structure factor calculations since the scattering effect of X-ray is reduced on the oscillating atoms compared to the atoms at rest [3]. The *B*-factor of the *j*th atom can be formally written as  $B_j = 8\pi^2 \langle r_j^2 \rangle$ , where  $r_j$  corresponds to the displacement of the *j*th atom from its equilibrium position. The thermal motion and mobility of protein atoms are closely linked to how the protein folds to the native state and how it interacts with specific binding partners in the cellular environment [4].

**Table 1.** Summary of the RSQ predictions on the 635 test proteins.

Groups	$N_{\text{pro}}$	TM	$d_o$ (Å)	$d_p$ (Å)	$\Delta d$ (Å)	PCC	AUC
C-score > -1.5	506	0.80	2.7 (0.79)	2.2 (0.82)	1.4 (0.12)	0.69	0.89
C-score < -1.5	129	0.40	10.8 (0.4)	8.6 (0.32)	6.4 (0.22)	0.53	0.78
Overall	635	0.71	4.3 (0.72)	3.4 (0.72)	2.4 (0.14)	0.66	0.87

Numbers in parentheses are the values computed after normalizing the distance  $d_p$  and  $d_o$  using Eq. (2).

$N_{\text{pro}}$ : Number of proteins in the set.

TM: Average TM-score of the first I-TASSER model.

$d_o$  ( $d_p$ ): Observed (predicted) distance between the model and the native structures.

$\Delta d$ : Average difference between  $d_o$  and  $d_p$ .

PCC: PCC between predicted and observed distances.

AUC: AUC of the ROC.

In this work, we aim to develop a unified pipeline, ResQ, for the estimation of the residue-specific quality (RSQ) of protein structure prediction and the inherent B-factor profile (BFP) of all residues along the chain. A major advantage of the pipeline over many other sequence or structure-based quality prediction approaches [4,5] is that ResQ takes into the consideration of the intermediate procedures of structure modeling simulations (e.g., structural variations of the simulation trajectories) and homology-based data mining (e.g., multiple template alignments from threading and structural alignment searches); the intermediate modeling features have also been used by several other structural error predictors [6,7]. Since these procedures are common setting in most of the cutting-edge protein structure assembly approaches, the inclusion of the intermediate features should not compromise the general usefulness of the ResQ program. The method also contains the single-model options to model RSQ and BFP from the structure alone<sup>†</sup>.

## Results and discussion

### Datasets

A total number of 1270 non-redundant single-domain proteins were collected from the PDB, which have a pair-wise sequence identity <30% with size from 50 to 300 residues. We randomly selected half of the proteins for training and the remaining for test of ResQ. Because the thermal motion factors in protein crystals can be affected by systematic errors such as experimental resolution, crystal contact and refinement procedures, the raw B-factor values are usually not comparable between different experimental structures. We calculate a normalized B-factor, which can reduce the systematical variations and is defined as:  $b_j = (B_j - \mu)/\sigma$ , where  $B_j$  is the raw B-factor value of the  $j$ th residue for the alpha-carbon atom;  $\mu$  and  $\sigma$  are the mean and standard deviation along the target sequence, respectively. The set of the normalized B-factor values along the target chain are called BFP.

### RSQ and BFP predictions by ResQ on I-TASSER models

**Overall RSQ results.** I-TASSER was used to generate structure predictions for 635 non-redundant testing proteins. A total of 483 proteins are categorized into Easy and 152 proteins are categorized into Hard targets according to the significance score of the LOMETS alignments [8]. After excluding the homologous templates, I-TASSER generated the first models with an average TM-score of 0.71 and an RMSD of 5.7 Å; these modeling results are largely consistent with the results from the I-TASSER modeling in the recent CASP experiments [9,10].

ResQ was applied to the first I-TASSER models to estimate the distance of each residue to the native. As shown in Table 1, the average distance by ResQ ( $d_p$ ) is 3.4 Å, which is consistently lower than the observed distance ( $d_o$ ) of the residues to the native (4.3 Å), resulting in an average difference between  $d_p$  and  $d_o$ ,  $\Delta d = 2.4$  Å. This consistent difference is mainly due to the lower distance estimation for the residues of large modeling errors. If we rescale  $d_p$  and  $d_o$  by the TM-score scale that was designed to depress contribution of large distance errors [11] [see Eq. (2) in Methods], the consistent difference between  $d_p$  and  $d_o$  disappears, both being 0.72 (see value in parentheses of Table 1).

We further split the test proteins into two groups based on the I-TASSER confidence score (C-score), that is, the high-confidence or low-confidence groups with a C-score above or below -1.5. As expected, models with a high C-score have a much better quality (TM-score = 0.8) than those with a low C-score (TM-score = 0.4). Accordingly, the RSQ prediction for the high C-score proteins is much more accurate ( $\Delta d = 1.4$  Å) than that of low C-score ( $\Delta d = 6.4$  Å). The average Pearson correlation coefficient (PCC) and area under the curve (AUC) values of the high C-score models are also higher than those of low C-score ones by 30% and 14%, respectively (Table 1).

The dependence of RSQ prediction on C-score is expected because for the targets of a low confidence, most of the I-TASSER models have an incorrect fold

(TM-score < 0.5) and the actual distance from model to the native is high (~10.8 Å). Thus, the estimation of such high distance is statistically more difficult. In Fig. 1a, we divide the residues into 20 bins according to their observed distance on the I-TASSER models to native ( $d_o$ ), and we compute the average errors of the predicted distances in each bin ( $\Delta d$ ). The majority of residues (>75%) was well modeled by I-TASSER with a distance to the native of <3 Å. ResQ predicted the distance for these residues accurately with a small  $\Delta d$ , that is, 0.56 Å, 0.60 Å and 0.96 Å for the distance bins [0, 1], [1, 2] and [2, 3], respectively. However, with the increase of the modeling error, the error of the RSQ predictions increases almost linearly with  $d_o$ .

**RSQ predictions on different local structures.** In Table S2 in the supplementary materials, we present ResQ predictions for the 506 targets with C-score > -1.5, with the residues split into six disjoint subsets: (*aligned* or *unaligned*) × (*alpha*, *beta* or *coil*). A residue is defined as *aligned* if there are >40% of LOMETS templates aligned on it, or *unaligned* otherwise, where *alpha*, *beta* and *coil* follow standard secondary structure definition.

Since I-TASSER modeling is built on multiple threading templates, there is a strong correlation between the structural error ( $d_o$ ) and the alignment coverage of the residues. The residues on the threading aligned regions have a much smaller distance deviation from the native (2.5 Å) than the unaligned residues (11.7 Å) because structures of on the threading aligned regions have a higher number of spatial restraints that result in a higher accuracy of modeling. Accordingly, the error of the estimated RSQ is lower in the threading aligned regions than those in the unaligned regions, as shown by the decreased  $\Delta d$  in Table S2 (Columns 2–7) mainly because the residues with smaller modeling errors tend to have a lower uncertainty and therefore ResQ can generate more accurate distance estimation on them (Fig. 1a). Similarly, structural modeling in the regular secondary structure regions (alpha and beta) is generally more accurate than that in the coil residues, which are true in both threading aligned and unaligned regions. The estimated RSQ is therefore closer to the actual modeling errors in these regions.

In Fig. 1b–d, we present an illustrative example from the PhoQ histidine kinase domain (PDB ID: 1id0A). In this example, the I-TASSER assembly simulations are more divergent in the loop region (D96-L111) as shown in Fig. 1c; this leads to a higher predicted distance deviation by ResQ that is consistent with the actual high modeling error of this region shown in Fig. 1b. The overall RSQ profile is in close agreement with the observed distance deviations as shown in Fig. 1d.

**BFP prediction.** Three approaches were tested to generate BFP predictions. The template-based prediction is generated by transferring the B-factors of the threading template proteins using Eq. (S3), while the profile-based prediction is by training the BFP data on the sequence profile generated by PSI-BLAST search. The third approach is a combinational method that trains the BFP on a combination of threading templates and sequence profiles. A summary of the PCC and AUC between the observed and predicted B-factors from the three approaches are listed in Table S3.

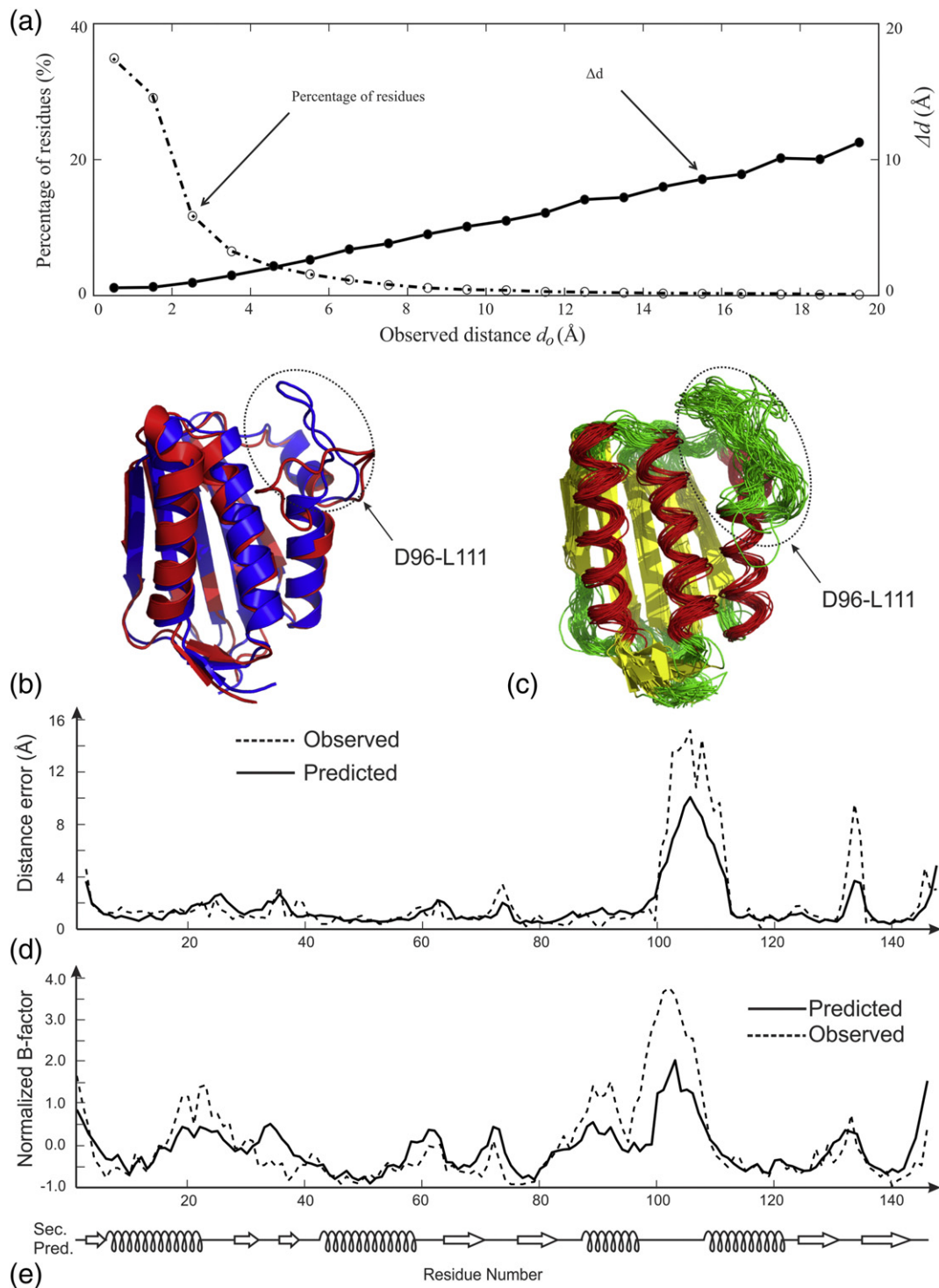
The profile-based training approach generated a slightly higher PCC value (0.59) than the template-based transferal (0.54), while the combination of the threading templates and sequence profiles achieves the highest PCC (0.61). The difference between the two methods (profile-based and combined) is statistically significant with the  $p$  value of the student  $t$ -test below  $10^{-12}$ . A similar tendency is followed by the AUC assessment, where the combined prediction outperforms both template-based and profile-based prediction methods.

In Fig. 1e, we show the predicted BFP for 1id0A by the combination-based approach. Interestingly, the highest B-factors occur mostly on the regions around the loop region D96-L111, which shows some level of correlation between local modeling error and BFP.

#### Comparison of ResQ predictions with other methods on the I-TASSER models

To benchmark ResQ with other methods, we downloaded and installed two recently published RSQ and BFP prediction programs to our computers. The SMOQ program was designed to predict residue-specific local quality based on machine learning, which has three options of basic (B), basic + profile (B + P) and basic + profile + SOV (B + P + S) [5]. The PROFbval program was developed for B-factor prediction trained on sequence profile and secondary structure predictions [4]. Both programs were run with default settings on the 635 testing proteins.

For the RSQ prediction, SMOQ is tested on the same set of the I-TASSER models. As shown in Table S4, the average error of the estimated distance deviation ( $\Delta d$ ) by ResQ is 2.4 Å, which is about 1.2 Å lower than the best results from the SMOQ program that uses the basic option (3.63 Å). If considering TM-score normalized distance,  $\Delta d$  of ResQ is also lower (0.14 versus 0.26). The average PCC of the RSQ predictions by ResQ is 47% higher than the best SMOQ results. The difference between ResQ and SMOQ is statistically significant with a  $p$  value <  $10^{-89}$  in the student  $t$ -test. One reason for the significant improvement of RSQ predictions might be that the ResQ has been trained on the I-TASSER-specific features, including structural variations of the templates and the simulation decoys



**Fig. 1.** RSQ and BFP predictions by ResQ. (a) Histogram and RSQ distribution at different modeling errors ( $d_o$ ) based on I-TASSER models from 506 test proteins with C-score  $> -1.5$ . Open circles connected by broken curve show the percentage of residues. Filled circles with continuous curve are the average prediction error ( $\Delta d$ ) in each distance bin. (b–e) An illustrative example from 1id0A showing the ResQ predictions. (b) Superposition of the X-ray structure (red) and the I-TASSER model (blue). (c) The ensemble of superimposed structure decoys in I-TASSER modeling. Alpha, beta and loop regions are shown in red, yellow and green, respectively. Structure regions with a high variation are highlighted by broken circle. (d) The predicted and observed distance errors. (e) The predicted and observed BFPs. The bottom panel shows secondary structure of the protein.

(see Table S1), which were not included in the SMOQ program. In fact, we have re-run ResQ by turning off the immediate features from I-TASSER simulations. The result of ResQ becomes much worse, which only marginally outperforms the best results from SMOQ (see Tables S1 and S4). Although the degradation is partly due to the fact that ResQ has been trained on the full set of sequence and structure features, this result indeed highlights the importance of the intermediate modeling features in model quality evaluations.

The last row in Table S4 lists the result of the *B*-factor prediction, where the PCC of ResQ (0.61) is 17% higher than that by PROFbval (0.52). Again, the difference is statistically significant, which corresponds to a *p* value in the student *t*-test below  $10^{-70}$ .

#### *Comparison of RSQ prediction with other methods on CASP decoys*

In this section, we test ResQ for the local structure quality prediction using the decoys that were generated in the recent CASP experiments, which gives us an opportunity to compare ResQ with many state-of-the-art model quality assessment programs (MQAPs). Because the structural decoys were generated by different methods from multiple laboratories, these testing data should also allow examining the robustness of the ResQ method, which was primarily trained on the I-TASSER decoys. Since the intermediate features from structure assembly simulation are not available on the CASP decoys, ResQ will skip these features to generate RSQ predictions. However, other features, including sequence-based and structured-based database searches, are still implemented.

The detailed comparisons of ResQ with the top CASP predictors are listed in Table S5 (for CASP9 decoys), Tables S6 and S7 (CASP10) and Tables S8 and S9 (CASP11), with analysis and discussion presented in Text S4. In general, local distance error prediction of ResQ consistently outperforms most of the CASP MQAP predictors, where the PCC and AUC scores by ResQ are among the top but often slightly outperformed by the best predictors (in particular, for AUC). The overall data suggest that the prediction results of the ResQ are comparable to or better than the state-of-the-art MQAP methods in CASP.

#### *Application of ResQ B-factor prediction to molecular replacement*

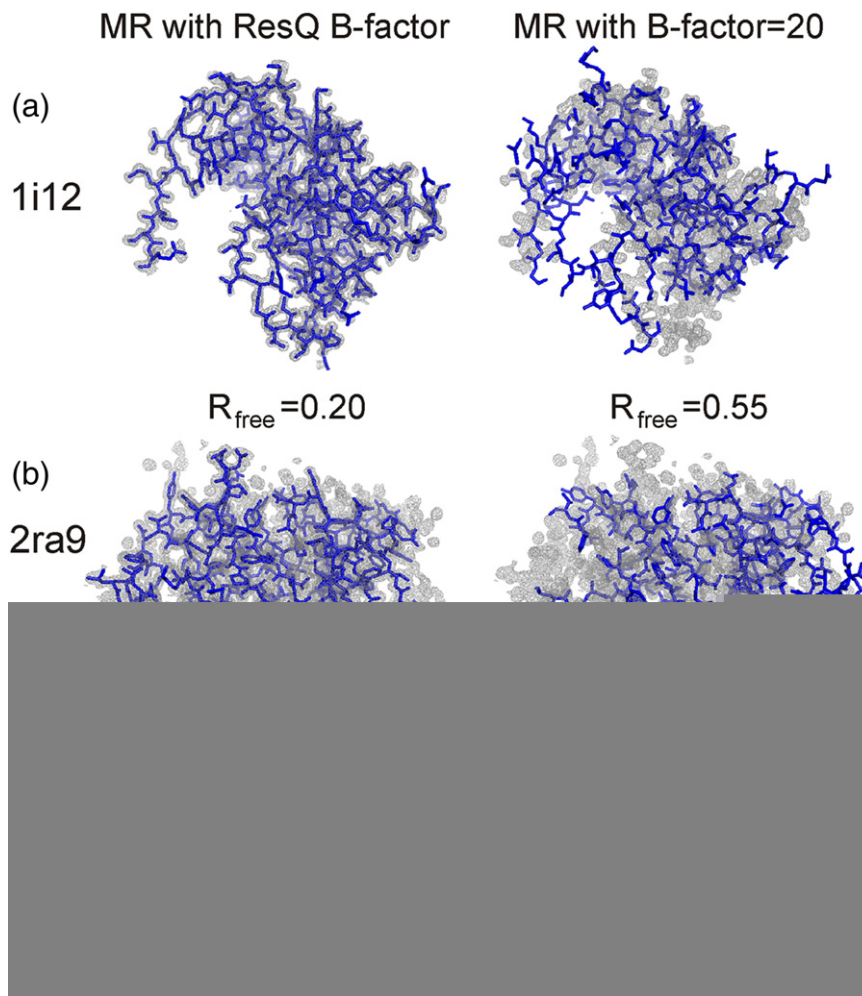
One of the important uses of the *B*-factor prediction is for assisting the molecular replacement (MR)-based structure determination in X-ray crystallography. In MR, close homologous structure models are used to replace the unknown targets for deciding the phase of the diffraction waves so that the electron density of the target protein can be calculated using

Fourier transformations from the diffraction pattern data. Recently, a closely related study suggests that local error estimates improve MR dramatically [12]. Different from the local error estimates that evaluate the quality of target models, thermal mobility of protein atoms can directly the intensity of the reflection waves. Therefore, appropriate *B*-factor estimations, which roughly reflect the atomic thermal mobility, should be important for correct structure factor calculation and MR solution.

To test the effect of ResQ *B*-factor prediction on MR, we collected 100 non-redundant proteins from the PDB, which have both X-ray structure and electron density data available and with fewer than 300 residues and  $\leq 4$  copies in the asymmetric unit. The I-TASSER models without using homologous templates are then used as the probe for MR. A progressive model truncation and editing procedure based on structural deviation score [13] was applied to generate up to 40 truncated models for each target, which are submitted to PHENIX [14] for automated phase determination and model reconstruction. It was shown that correct MR solutions could be obtained for 54 out of the 100 proteins if we use an optimal constant *B*-factor (20). Here, an MR solution is defined as correct if more than 25% of the target sequence can be built by the automated MR procedure with the final structure models being closer to the experimental structure than the starting model.

When we applied the predicted *B*-factor by ResQ to the MR program, three additional proteins (PDB IDs: 1i12, 2ra9 and 2tnf) have the MR solution successfully obtained. Figure 2 shows a comparison of the final models for the three proteins overlaid on the experimental electron density maps, which were obtained using the ResQ *B*-factor prediction and the optimal constant *B*-factor, respectively. Even though the same I-TASSER model and the model editing procedure were applied, the use of the ResQ *B*-factor prediction resulted in much closer fits of the final models with the electron density maps. The average  $R_{\text{free}}$  value, which measures how well the simulated diffraction data match with the experimentally observed diffraction pattern, was significantly reduced from 0.53 to 0.27.

We note that the current strategy, that is, using RSQ to truncate structural models and then using BFP to scale atomic motion, generated the best MR results with 57 successful targets in our test. Simply replacing *B*-factor by RSQ does not improve the results, probably due to the fact that the RSQ information was already used in the model truncation procedure. We also tested the procedure with using RSQ as *B*-factor on the full-length models without truncation, which could result in slightly better MR results than that using constant *B*-factor; however, the overall results are much worse than the optimal results with model truncations [13].



**Fig. 2.** Structure models generated by MR overlaid on the experimental electron density image. The left and right panels are MR results using the ResQ predicted and constant *B*-factors, respectively. (a) 2ra9; (b) 1i12; (c) 2tnf.

## Conclusion

We developed a new algorithm, ResQ, for unified predictions of protein RSQ and BFP. One of the major advantages of the method is the integration of the intermediate information of structure modeling, including threading coverage and conformational variations, with the structure profile information from homologous database searches.

ResQ was first tested on a set of 635 non-homologous proteins with structural models by I-TASSER simulations. The residue-level distance to the native could be predicted with an average error of 2.4 Å. For the models with C-score > -1.5, which have a better global quality and are thus more relevant to biological uses, the distance error of RSQ is reduced to 1.4 Å. Detailed data analyses showed that the absolute error of RSQ is highly correlated with the quality of the structure models, where the RSQ has a much higher accuracy in the conserved regions of regular secondary structures

than that in other threading unaligned and coil/tail regions. The overall results of ResQ on both RSQ and *B*-factor predictions showed advantage over other methods that do not use the intermediate structural features.

Second, we tested ResQ on the CASP9-11 decoys generated by various structure prediction methods. Although no intermediate structure features are used, the single-model mode prediction of ResQ, built on the sequence-based and structure-based homologous database searches, generated local structural quality estimations with an accuracy comparable (or superior) to the best-performing MQAP methods that have been trained from various consensus features and/or statistics potentials. The data demonstrate the robustness of the ResQ algorithm to deal with both cases with small and large number of structure decoys. However, when the number of reference decoys decreased, ResQ's performance was slightly degraded, as seen by the difference in performance on the CASP Stage 1 and 2 decoys. Part of the degradation

is due to the reduction of the overall decoys in Stage 1; this suggests the room for further improving ResQ, for example, by exploring multiple statistical potentials [15], in particular, for the low-quality decoys.

As a test of the usefulness of the ResQ predictions, the predicted BFP was applied to 100 medium-size to large-size proteins for assisting MR in X-ray crystallography. Starting from the same set of the I-TASSER models, the use of the ResQ predicted  $B$ -factor resulted in successful MR solutions on three proteins (PDB IDs: 1i12, 2ra9 and 2tnf) that were not able to solve from the constant  $B$ -factor setting, whereby the average  $R_{\text{free}}$  value was dramatically reduced from 0.53 to 0.27 in the three examples.

The ResQ algorithm has been successfully integrated with the I-TASSER pipeline, including online server and standalone package, for the RSQ and BFP estimations. Although ResQ was primarily trained with the I-TASSER decoys, the program can be used for models generated by other structure prediction methods since the intermediate features, including query-to-template alignment and structure decoy clustering, are standard output in the pipelines of most state-of-the-art structure prediction approaches. As an illustration, we list in Table S10 the application of ResQ to the models generated by QUARK-based *ab initio* simulation [16] on a set of 50 non-redundant proteins. Although the average RSQ error is higher than that for the I-TASSER models due to the relatively lower accuracy of the global models, for the top 20 proteins, most having a TM-score above 0.5, the average RSQ error is 1.6 Å for the QUARK models, comparable to the results for the confidently predicted I-TASSER models.

## Methods

### ResQ method

Support vector regressions, with the implementation provided in the SVM-light package<sup>‡</sup>, are used to train the RSQ and BFP data, with parameters and combinations of features optimized on the 635 training proteins. The I-TASSER simulations [17] are conducted to generate structure predictions for the training proteins, where homologous templates have been excluded from the threading template library. A brief outline of the I-TASSER pipeline is described in Text S1 in the supplementary materials.

For RSQ prediction, a total of 12 features are used to represent each residue: 3 from structural assembly simulations [i.e., average variation and standard deviation in Eq. (S1) and relative size of SPICKER cluster], 1 from threading templates distance variation [Eq. (S2)], 1 from TM-align structure alignment templates distance variation, 2 from threading alignment coverage computed from the top 200 and 10 templates, respectively, and 5 from the consistency between model and sequence-based feature predic-

tions (i.e., three probabilities in alpha-helix, beta-strand and coil states; one for secondary structure state; and one for relative solvent accessibility). A detailed description of the RSQ features and their respective contributions to the predictions are provided in Text S2 and Table S1.

For the BFP prediction, by using the window size of 9, a total of 72 ( $8 \times 9$ ) features are used to represent each residue. For each residue, the 8 features include 2 from template-based assignments based on LOMETS and TM-align, respectively [see Eq. (S3)]; 1 from alignment coverage with the top 200 LOMETS templates, 2 from relative solvent accessibility and secondary structure predictions and 3 from secondary structure profile including probabilities in alpha-helix, beta-strand and random-coil states (see Text S3).

After trials and errors on the training dataset, we found that the best results are generated using the linear support vector regressions in SVM-light with the parameters “-z r -c 0.5 -w 0.5”, for both RSQ and BFP predictions.

Although ResQ has been designed to generate both RSQ and BFP prediction, we note that RSQ and BFP are two distinct concepts, as RSQ is closely related with model construction methods while  $B$ -factor reflects inherent motions of atoms in the physiological environment that are independent from the modeling process. In fact, we collected the data of all the 1270 training and testing proteins and found that the PCC between  $B$ -factor and the local error of I-TASSER models is only 0.38<sup>§</sup>. This weak correlation is probably due to the fact that the sequence regions with a high  $B$ -factor are often less conserved and thus models generated by template-based methods (such as I-TASSER) tend to have a higher local error on these regions. However, a determinate relation between these two qualities does not exist. Nevertheless, a combined prediction on RSQ and BFP is feasible since both request training on multiple local sequence and structure features. Such combination should help make ResQ convenient for multiple uses on, for example, local accuracy estimation and MR.

### Assessment criteria of the RSQ and BFP predictions

Three measures are used to evaluate the accuracy of the RSQ prediction, where each measure is first computed per model and then presented as an average over all models. The first measure is the average difference ( $\Delta d$ ) between the predicted ( $d_p$ ) and observed ( $d_o$ ) distances of the model relative to the native, that is,

$$\Delta d = \frac{1}{L} \sum_{i=1}^L |d_o^i - d_p^i| \quad (1)$$

where  $L$  is the length of the protein and the distances are calculated based on the TM-score superposition

[11]. One flaw of this metric is that  $\Delta d$  can be dominated by the large distance pairs, where distinguishing distance errors beyond some cutoffs can be meaningless (e.g., an error of 10 Å is not necessarily better than 20 Å). Thus, we also use a metric based on a normalized distance to depress the contribution of large distance errors in RSQ evaluation, that is,

$$\Delta d' = \frac{1}{L} \sum_{i=1}^L |d_{o'}^i - d_{p'}^i| \quad (2)$$

where  $d_{o,p}' = 1/[1 + (d_{o,p}/d_s)^2]$  and  $d_s = 1.24\sqrt[3]{L-15} - 1.8$  are a TM-score-like scale to rule out protein length dependence of the distances [11].

The second measure is the PCC between  $d_p$  and  $d_o$ . The third measure is the AUC of the receiver-operating characteristic (ROC), designed to evaluate the ability of ResQ in discriminating between well and badly modeled regions, where a residue is defined as “well modeled” (positive) if the distance from model to the native is  $<3.8$  Å upon the TM-score superposition, otherwise as “badly modeled” (negative). PCC and AUC are the same as used by the CASP assessors for evaluating the accuracy of model quality estimation [18]. Following Kryshchuk *et al.*, we converted  $d_p$  into the range of (0, 1) by  $d_{pn} = 1/[1 + (d_p/5)^2]$  in the AUC calculation so that a fixed number of divisions can be used for different data samples to draw the ROC curves. Here, the selection of 3.8 Å as a cutoff is from the CASP assessors, which may be too high in this study since the AUC values are generally high for most predictions. We tested the impact of the distance cutoff on the AUC calculations using 635 test proteins. As expected, the AUC value decreases slightly with reduced distance cutoff, that is, the average AUC decreases from 0.88 to 0.87, 0.86, 0.84 and 0.79, respectively, when the distance cutoff reduced from 5 to 4, 3, 2 and 1 Å.

The BFP prediction is evaluated by the Pearson's correlation between the predicted and the experimental  $B$ -factors, which was also used in previous  $B$ -factor prediction studies [4]. Similar to the RSQ evaluation, we also use AUC for measuring the ability in discriminating between stable and flexible residues in structures, where a residue is defined as stable (positive) if the normalized  $B$ -factor is below 0 or as flexible (negative) otherwise. For uniform ROC division, we have normalized the predicted  $B$ -factor values ( $b$ ) to the range of (0, 1) by  $1/[1 + \exp(-b)]$ .

## Acknowledgements

We are grateful to Baoji He for providing the QUARK models. The project is supported in part by

the National Institute of General Medical Sciences (GM083107 and GM084222), National Natural Science Foundation of China (NSFC) Grant (11501306), and Nankai University Start-up Grant (ZB15006103).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jmb.2015.09.024>.

Received 23 June 2015;

Received in revised form 23 August 2015;

Accepted 28 September 2015

Available online 3 October 2015

### Keywords:

protein structure prediction;  
model quality evaluation;  
local error estimation;  
 $B$ -factor profile

†An online server and the source code of the ResQ program are freely available at <http://zhanglab.ccmb.med.umich.edu/ResQ/>.

‡<http://svmlight.joachims.org/>.

§[http://zhanglab.ccmb.med.umich.edu/ResQ/benchmark/RB\\_cc.txt](http://zhanglab.ccmb.med.umich.edu/ResQ/benchmark/RB_cc.txt).

### Abbreviations used:

RSQ, residue-specific quality; BFP,  $B$ -factor profile; PCC, Pearson correlation coefficient; AUC, area under the curve; MQAP, model quality assessment program; MR, molecular replacement; ROC, receiver-operating characteristic.

## References

- [1] J. Moult, A decade of CASP: Progress, bottlenecks and prognosis in protein structure prediction, *Curr. Opin. Struct. Biol.* 15 (2005) 285–289.
- [2] Y. Zhang, Protein structure prediction: When is it useful? *Curr. Opin. Struct. Biol.* 19 (2009) 145–155.
- [3] D. Sherwood, Cooper J. Crystals, X-rays and Proteins: Comprehensive Protein Crystallography, Oxford University Press, Oxford, UK, 2011.
- [4] A. Schlessinger, G. Yachdav, B. Rost, PROFbval: Predict flexible and rigid residues in proteins, *Bioinformatics* 22 (2006) 891–893.
- [5] R. Cao, Z. Wang, Y. Wang, J. Cheng, SMOQ: A tool for predicting the absolute residue-specific quality of a single protein model with support vector machines, *BMC Bioinf.* 15 (2014) 120.
- [6] B. Wallner, ProQM-resample: Improved model quality assessment for membrane proteins by limited conformational sampling, *Bioinformatics* 30 (2014) 2221–2223.
- [7] L.J. McGuffin, M.T. Buenavista, D.B. Roche, The ModFOLD4 server for the quality assessment of 3D protein models, *Nucleic Acids Res.* 41 (2013) W368–W372.



- [8] S. Wu, Y. Zhang, LOMETS: A local meta-threading-server for protein structure prediction, *Nucleic Acids Res.* 35 (2007) 3375–3382.
- [9] Y. Zhang, Interplay of I-TASSER and QUARK for template-based and *ab initio* protein structure prediction in CASP10, *Proteins* 82 (2014) 175–187.
- [10] D. Xu, J. Zhang, A. Roy, Y. Zhang, Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based *ab initio* folding and FG-MD-based structure refinement, *Proteins* 79 (2011) 147–160.
- [11] Y. Zhang, J. Skolnick, Scoring function for automated assessment of protein structure template quality, *Proteins* 57 (2004) 702–710.
- [12] G. Bunkoczi, B. Wallner, R.J. Read, Local error estimates dramatically improve the utility of homology models for solving crystal structures by molecular replacement, *Structure* 23 (2015) 397–406.
- [13] Y. Wang, Z. Xue, J. Virtanen, Y. Zhang, Molecular replacement for distant-homology proteins using iterative fragment assembly and progressive sequence truncation, 2015 (submitted for publication).
- [14] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, et al., PHENIX: A comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* 66 (2010) 213–221.
- [15] P. Benkert, S.C. Tosatto, D. Schomburg, QMEAN: A comprehensive scoring function for model quality assessment, *Proteins* 71 (2008) 261–277.
- [16] D. Xu, Y. Zhang, *Ab initio* protein structure assembly using continuous structure fragments and optimized knowledge-based force field, *Proteins* 80 (2012) 1715–1735.
- [17] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: Protein structure and function prediction, *Nat. Methods* 12 (2015) 7–8.
- [18] A. Kryshtafovych, A. Barbato, K. Fidelis, B. Monastyrskyy, T. Schwede, A. Tramontano, Assessment of the assessment: Evaluation of the model quality estimates in CASP10, *Proteins* 82 (2014) 112–126.