J Biomol NMR (2015) 62:511–525
DOI 10.1007/s10858-015-9914-y

ARTICLE

NMR data-driven structure determination using NMR-I-TASSER in the CASD-NMR experiment

Richard Jang · Yan Wang · Zhidong Xue · Yang Zhang

Received: 28 December 2014 / Accepted: 21 February 2015 / Published online: 4 March 2015

Abstract NMR-I-TASSER, an adaption of the I-TASSER algorithm combining NMR data for protein structure determination, recently joined the second round of the CASD-NMR experiment. Unlike many molecular dynamics-based methods, NMR-I-TASSER takes a molecular replacement-like approach to the problem by first threading the target through the PDB to identify structural templates which are then used for iterative NOE assignments and fragment structure assembly refinements. The employment of multiple templates allows NMR-I-TASSER to sample different topologies while convergence to a single structure is not required. Retroactive and blind tests of the CASD-NMR targets from Rounds 1 and 2 demonstrate that even without using NOE peak lists I-TASSER can generate correct structure topology with 15 of 20 targets having a TM-score above 0.5. With the addition of NOE-based distance restraints, NMR-I-TASSER significantly improved the I-TASSER models with all models having the TM-score above 0.5. The average RMSD was reduced from 5.29 to 2.14 Å in Round 1 and 3.18 to 1.71 Å in Round 2. There is no obvious difference in the modeling results with using raw and refined peak lists, indicating robustness of the pipeline to the NOE assignment errors. Overall, despite the low-resolution modeling the current NMR-I-TASSER pipeline provides a coarse-grained structure folding approach complementary to traditional molecular dynamics simulations, which can produce fast near-native frameworks for atomic-level structural refinement.

Keywords NMR restraints · NOE assignment · CASD-NMR · I-TASSER · Protein structure prediction

R. Jang · Z. Xue (✉)
School of Software Engineering, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China
e-mail: zdxue@hust.edu.cn

R. Jang · Y. Wang · Z. Xue · Y. Zhang (✉)
Department of Computational Medicine and Bioinformatics, University of Michigan, 100 Washtenaw Avenue, Ann Arbor, MI 48109-2218, USA
e-mail: zhng@umich.edu

Y. Wang
School of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

Y. Zhang
Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109, USA

Introduction

Methods for automated protein structure determination by nuclear magnetic resonance (NMR) spectroscopy have matured to the point that they are used by structural genomics centers, such as the Northeast Structural Genomics Consortium (NESG; http://www.nesg.org), for large-scale protein structure determination. Nevertheless, the use of automated methods is still far from being routine, especially for large proteins and cases with sparse or noisy data. To overcome the limitations of traditional molecular dynamics-based NMR methods, approaches from the protein structure prediction community, which are designed to predict 3D structures from amino acid sequence only, have been adapted for NMR (Li et al. 2004; Meiler and Baker 2003; Montelione 2014). These methods take advantage of restraints derived from homologous structures from the Protein Data Bank (PDB), which can implement the restraints derived from NMR experiments, so that the structure sampling has more restraints to lead to convergence.

Meanwhile, traditional NMR and structure prediction methods share the same underlying computational problem of structure sampling to minimize the energy functions. One example of the advanced structure prediction methods is I-TASSER (Roy et al. 2010; Yang et al. 2015), which was designed to identify homologous and analogous templates from the PDB by meta-threading; the full-length models are then constructed by reassembling the continuously aligned fragments excised from the threading templates, under the guidance of the consensus contact/distance restraints collected from multiple threading alignments. To examine the capacity of I-TASSER in NMR data-driven structure determination, we have developed NMR-I-TASSER, which combines the fragment assembly simulations with distance restraints from NOE assignment for template-based and reduced-level protein structure determination.

A preliminary version of NMR-I-TASSER recently joined the Round 2 of the Critical Assessment of Automated Structure Determination of Proteins from NMR Data (CASD-NMR) experiment. This manuscript will describe the blind testing results, as well as the data of applying NMR-I-TASSER retroactively to the other Round 2 targets and all Round 1 targets. To ensure an objective evaluation, we have excluded threading templates whose PDB deposition date occurs after the peak list release date of the CASD-NMR targets. Although NMR-I-TASSER was a late entrant to CASD-NMR, its methods can provide a different perspective to the NMR structure determination problem. Our methodology is based on a molecular replacement approach to NMR, in the way that homologous structures are first identified by threading and NOE assignment is then performed from these structures to provide restraints for the structure sampling step. Molecular replacement approaches for NMR based on threading is not a new concept. For example, CS23D (Wishart et al. 2008) uses chemical shift threading to identify templates. Nevertheless, the threading-based approaches for NMR have not been explored as widely as the molecular dynamics approaches.

We will first examine how I-TASSER alone without NMR data performs on the CASD-NMR targets and then show that incorporating restraints from NOE peak lists can consistently improve the structure modeling. In situations where no suitable templates exist, normal structure assembly simulations can often get stuck in incorrect conformations. We show here that even when no suitable template exists, the restraints from NOEs can still guide I-TASSER towards the correct topology. Finally, a comparison of the NOE assignment accuracies and structure results between raw and refined peak lists will be presented to examine how sensitive the NMR-I-TASSER modeling results are to the errors found in the raw peak lists.

Methods

NMR-I-TASSER is an adaptation of the I-TASSER protocol (Roy et al. 2010; Yang et al. 2015) for NOE peak lists. It consists of 4 main steps that form an iterative pipeline: template identification, NOE assignment, structure sampling, and model selection by structure clustering. The pipeline is illustrated in Fig. 1. In general, the multiple threading approach LOMETS (Wu and Zhang 2007) was used to identify homologous structures, and then the template alignments were used as input to guide the NOE assignment. I-TASSER based fragment assembly simulations are then used for conformational sampling, with the lowest free energy conformations identified by structure clustering (Zhang and Skolnick 2004b) which are used for further NOE assignment and structure filtering. The process was iteratively repeated several times until convergence (Fig. 1). For a protein of 150 residues long, the typical CPU time cost by NMR-I-TASSER is around 10-15 h on a 2.8 GHz IBM NeXtScale n1200 machine, with the majority of the time spent on the I-TASSER based structure assembly simulations.

Template identification

The first step of NMR-I-TASSER consists of searching the target sequence through the PDB for homologous/analogous structures by threading (Bowie et al. 1991). In contrast to traditional homology search methods based only on sequence comparison, such as PSI-BLAST (Altschul et al. 1997), threading is often more sensitive. In addition to the sequence profiles derived from PSI-BLAST (Altschul et al. 1997) or hidden Markov models (Eddy 1998), structural features including secondary structure prediction, torsion angle and solvent accessibility are often exploited in the threading approaches (Wu and Zhang 2008). This enables the identification of templates that have similar fold to the target but low sequence similarity. The threading step also results in a sequence-structure alignment that gives the conserved structure fragments in the aligned regions. These fragments can then be used in the structure assembly step. In NMR-I-TASSER, the program LOMETS was used (Wu and Zhang 2007), which is meta-server consisting of 8 complementary threading programs. LOMETS calculates a confidence score to predict whether the threading results are Easy, Medium, or Hard based on the Z-scores of each program, while an Easy target with an higher average Z-score often have template structure with a higher accuracy.

The threading templates are filtered by the overlap score of the contact map and the NOE data. Before the filtering procedure can be conducted, the missing atoms in the templates need to be reconstructed since the sequence-

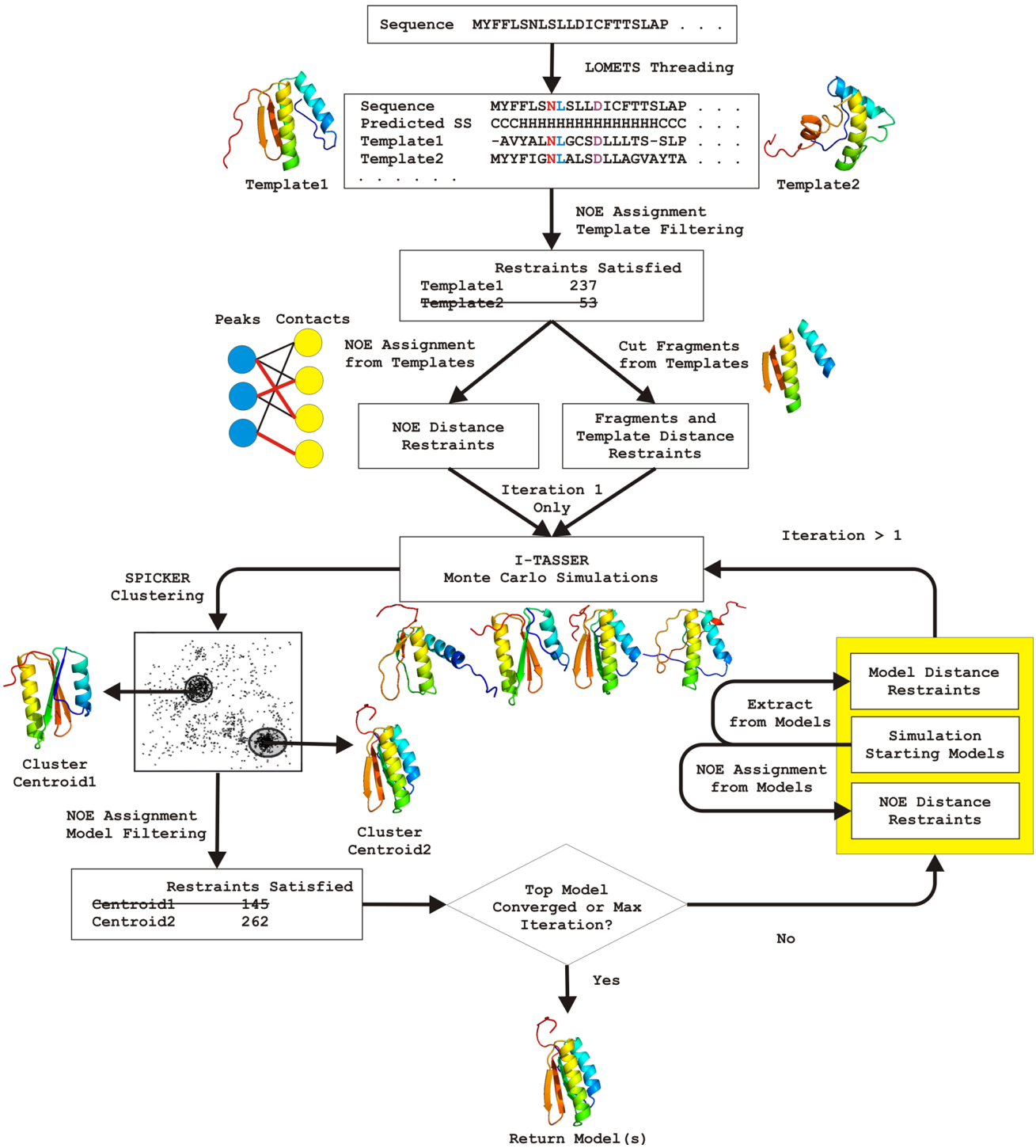

Fig. 1 Flowchart of the NMR-I-TASSER pipeline for iterative NMR data-driven structure determination and refinement. After the initial step of LOMETS threading, each iteration consists of structure-guided NOE assignment, I-TASSER-based structure assembly simulations, and SPICKER decoy clustering and model selection

structure alignment may have gaps and insertions. FG-MD (Zhang et al. 2011) was used to add the missing heavy atoms from the threading alignment, and HAAD (Li et al. 2009) was used to add the protons. Here, FG-MD is a simulated-annealing based molecular dynamics (MD) simulation algorithm developed for atomic-level protein structure refinement and reconstruction (normally from Cα-trace models). One difference from normal MD simulation methods is that FG-MD collects restraints from analogous structure fragments detected from the PDB, which are used

to improve the funnel shape of energy landscape of the physics-based force field to improve the quality of the final atomic models (Zhang et al. 2011). HAAD is a program to quickly add protons to a heavy-atom model by optimizing the hydrogen-bonding network and minimizing steric clashes (Li et al. 2009). The NOE assignment algorithm was run separately for each template model as input, and then the models were ranked by contact overlap. The top 10 templates were used as input for another NOE assignment step to obtain the distance restraints for structure sampling. The top 40 templates were used as starting models for the Monte Carlo simulations in the structural sampling step.

NOE assignment

One of the key steps in the NMR-I-TASSER pipeline is NOE assignment because it is used in multiple steps of the procedure, including structure sample simulation and model selection. As shown in Fig. 2, the NOE assignment procedure consists of the following steps: (1) generating assignment possibilities, (2) scoring each possibility, (3) filtering by assignment score and structure, (4) generating the one-to-one seed assignments, (5) calibrating the distances, and (6) outputting the final restraints for I-TASSER structure sampling.

The one-to-one seed assignments represent the more reliable assignments and the ambiguous assignments represent the less reliable assignments (Nilges and O'Donoghue 1998). Each ambiguous assignment consists of multiple assignment possibilities for a single NOE peak. An ambiguous assignment is satisfied if at least one of its assignment possibilities is satisfied. To handle the lower reliability of ambiguous assignments, the ambiguous assignments that have less than 3 possibilities were grouped together.

Generating assignment possibilities

We first describe the input of the NOE assignment procedure and introduce the necessary notations. The input data consists of (1) a set S of 3D model structures obtained from

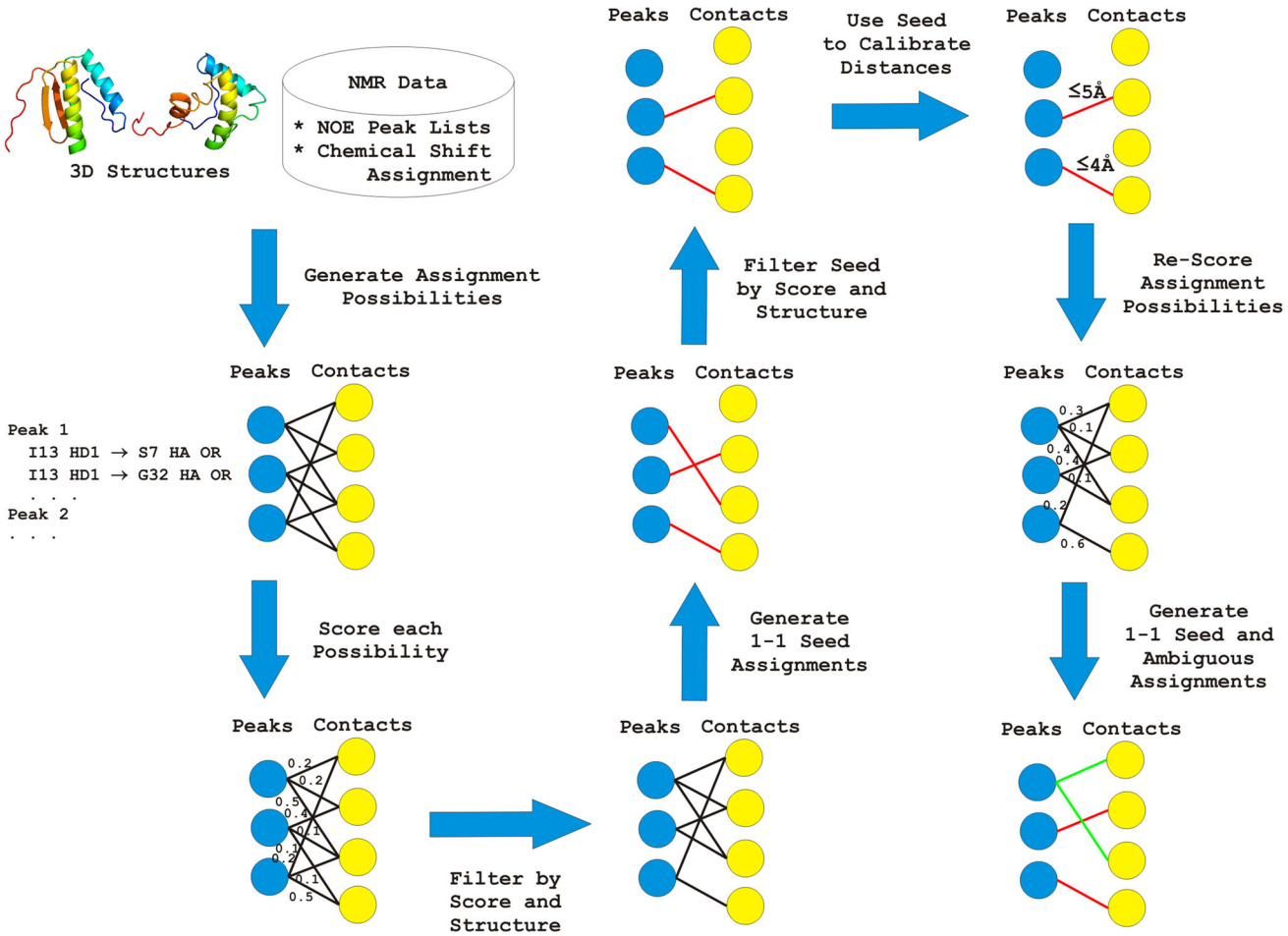

Fig. 2 Overview of the NOE assignment pipeline used in NMR-I-TASSER. The input consists of the chemical shift assignment list, NOE peak list, and 3D structures

either the LOMETS threading or I-TASSER clusters; (2) chemical shift assignment list; and (3) 3D NOE peak lists. Since 3D NOE peaks were used, a potential contact $\boldsymbol{c}$ between atoms hx_c and h_c can be represented as a 3-tuple $\boldsymbol{c} = (x_c, hx_c, h_c)$, where x_c is a heavy atom (C or N), hx_c is the proton covalently attached to x_c, and h_c is a proton believed to be in contact with hx_c. For notational purposes, R_{hx_c} shall represent the residue number of the amino acid containing proton hx_c, and R_{h_c} shall represent the residue number of the amino acid containing proton h_c. The distance of contact $\boldsymbol{c}$ in structure $s \in \boldsymbol{S}$ shall be denoted as $d_{\boldsymbol{c},s}$, and the minimum distance of $\boldsymbol{c}$ among all structures in $\boldsymbol{S}$ shall be denoted as $d^{\min}_{\boldsymbol{c},\boldsymbol{S}}$.

Denoting the chemical shift assignment of atom k as δ_k, the chemical shifts of each contact $\boldsymbol{c} = (x_c, hx_c, h_c)$ can be represented as the 3-tuple $\boldsymbol{\delta_c} = (\delta_{xc}, \delta_{hxc}, \delta_{hc})$. Each 3D peak $\boldsymbol{p}$ can be represented as a 4-tuple of the form $\boldsymbol{p} = (\delta_{p_1}, \delta_{p_2}, \delta_{p_3}, I_p) = (\boldsymbol{\delta_p}, I_p)$ where δ_{p_1} is the chemical shift of some unknown heavy atoms, δ_{p_2} is the chemical shift of the protons covalently attached to the heavy atoms, and δ_{p_3} is the chemical shift of some unknown protons whose upper bound distance to the proton with chemical shift δ_{p_2} is given by peak intensity I_p. The NOE assignment problem is to determine the identity of the unknown atoms for each peak, to extract an upper bound distance from the intensity, and to decide whether the peak represents a genuine contact or is noise.

Given the chemical shift assignments and chemical shift match tolerances $\boldsymbol{\Delta} = (\Delta\delta_1, \Delta\delta_2, \Delta\delta_3)$ for each peak dimension, the set of all assignment possibilities $\boldsymbol{A_p}$ for peak $\boldsymbol{p}$ can be enumerated as the set of all contacts along with their upper distance bounds

$$\boldsymbol{A_p} = \left\{ (\boldsymbol{c}, d_p) \middle| \left|\boldsymbol{\delta_p} - \boldsymbol{\delta_c}\right| < \boldsymbol{\Delta}, d_p = {}^{k}\!/_{I_p^6} \right\} \tag{1}$$

assuming the upper bound distance d_p is given by the isolated spin pair approximation, where k is a constant determined in the distance calibration step. For this work, we used $\boldsymbol{\Delta} = (0.35, 0.035, 0.035)$. Finally, the set of all assignment possibilities $\boldsymbol{A}$ for all peaks can be expressed as the set of peak-contact pairs

$$\boldsymbol{A} = \left\{ (\boldsymbol{p}, \boldsymbol{c}) | \boldsymbol{c} \in \boldsymbol{A_p} \right\} \tag{2}$$

The assignment problem can be modeled as assigning at most one contact to each peak. Contacts to methyl protons were treated as a group of contacts consisting of separate contacts to each of the 3 protons in the methyl group.

Scoring assignment possibilities

Each assignment possibility $(\boldsymbol{p}, \boldsymbol{c})$ is given a score

$$\begin{aligned} Score(\boldsymbol{p}, \boldsymbol{c} | \boldsymbol{A}, \boldsymbol{S}) &= S_{cs}(\boldsymbol{\delta_p}, \boldsymbol{\delta_c}) + S_{str}(\boldsymbol{c}, \boldsymbol{S}) \\ &+ S_{int}(d_p, d^{\min}_{\boldsymbol{c},\boldsymbol{S}}) + S_{sym}(\boldsymbol{c}, \boldsymbol{A}) \\ &+ S_{inter}(\boldsymbol{c}, \boldsymbol{A}) + S_{net}(\boldsymbol{c}, \boldsymbol{A}) + S_{netstr}(\boldsymbol{c}, \boldsymbol{A}, \boldsymbol{S}) + \theta(\boldsymbol{c}) \end{aligned} \tag{3}$$

The score takes into account the contacts found in predicted structures and also other assignment possibilities. The concepts used are similar to the terms found in the scoring function of CANDID (Herrmann et al. 2002), such as the closeness of the chemical shift fit, the presence of symmetry-related cross-peaks, network anchoring, and compatibility with intermediate 3D structures.

In Eq. (3), $S_{cs}(\boldsymbol{\delta_p}, \boldsymbol{\delta_c})$ gives the closeness of the chemical shift fit between the chemical shifts of a peak and the chemical shifts of the atoms in a contact, which is defined as

$$S_{cs}(\boldsymbol{\delta_p}, \boldsymbol{\delta_c}) = \sum_{i=1}^{3} f_{erfc}(|\delta_{p_i} - \delta_{c_i}|, 0, \sigma_i) \tag{4}$$

where f_{erfc} is the Gaussian complementary error function with mean 0 and standard deviation σ_i. σ_i was set such that a chemical shift difference of $\Delta\delta_i$ gives a value of 0.05.

$S_{str}(\boldsymbol{c}, \boldsymbol{S})$ gives the compatibility of the contact with the input 3D structures, which is defined as

$$S_{str}(\boldsymbol{c}, \boldsymbol{S}) = \frac{M(\boldsymbol{c}, \boldsymbol{S})}{M(\boldsymbol{S})} \tag{5}$$

where $M(\boldsymbol{c}, \boldsymbol{S})$ is the number of structures with $d^{\min}_{\boldsymbol{c},\boldsymbol{S}} \leq T_{cut}$, and $M(\boldsymbol{S})$ is the total number of structures. For this work, T_{cut} was set to 6 Å.

$S_{int}(d_p, d^{\min}_{\boldsymbol{c},\boldsymbol{S}})$ gives the similarity between the distance of the contact in the 3D structures and the calibrated upper bound distance from the peak intensity, which is defined as

$$S_{int}\left(d_p, d^{\min}_{c,S}\right) = \begin{cases} 1 & \text{if } d^{\min}_{c,S} \leqslant d_p \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$S_{sym}(\boldsymbol{c}, \boldsymbol{A})$ considers whether or not there exists an assignment possibility for the symmetric contact (y_c, h_c, hx_c) of contact $\boldsymbol{c}$, where y_c is the heavy atom covalently bound to h_c. It is defined as

$$S_{sym}(\boldsymbol{c}, \boldsymbol{A}) = \begin{cases} 1 & \text{if exists } (y_c, h_c, hx_c) \in \boldsymbol{A} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$S_{inter}(\boldsymbol{c}, \boldsymbol{A})$ measures the assignment possibilities between the residues R_{hx_c} and R_{h_c}, which is defined as

$$S_{inter}(\boldsymbol{c}, \boldsymbol{A}) = \frac{Q(R_{hx_c}, R_{h_c}, \boldsymbol{A})}{Q(R_{hx_c}, R_{h_c})} \tag{8}$$

where $Q(R_{hx_c}, R_{h_c}, \boldsymbol{A})$ is the number of proton pairs between residues R_{hx_c} and R_{h_c} that have at least one assignment possibility in $\boldsymbol{A}$, and $Q(R_{hx_c}, R_{h_c})$ is the number of proton pairs between R_{hx_c} and R_{h_c}.

$S_{net}(\boldsymbol{c},\boldsymbol{A})$ is similar to $S_{inter}(\boldsymbol{c},\boldsymbol{A})$ except that it considers the number of assignment possibilities between the protons of neighboring residues defined by a sequence window, which is defined as

$$S_{net}(\boldsymbol{c},\boldsymbol{A}) = \frac{T(R_{hx_c}, R_{h_c}, W, \boldsymbol{A})}{T(R_{hx_c}, R_{h_c}, W)} \tag{9}$$

where $T(R_{hx_c}, R_{h_c}, W, \boldsymbol{A})$ is the number of proton pairs between the neighboring residues R_{hx_c} and R_{h_c} that have at least one assignment possibility in $\boldsymbol{A}$, and $T(R_{hx_c}, R_{h_c}, W)$ is the number of such proton pairs ignoring whether or not they have an assignment possibility. W is the sequence window, and it was set to 1 to consider the residues sequentially adjacent to R_{hx_c} and R_{h_c}.

$S_{netstr}(\boldsymbol{c},\boldsymbol{A},\boldsymbol{S})$ is similar to $S_{net}(\boldsymbol{c},\boldsymbol{A})$ except that it considers whether or not the neighboring contacts are also in the input structures, i.e.

$$S_{net}(\boldsymbol{c},\boldsymbol{A},\boldsymbol{S}) = \frac{V(R_{hx_c}, R_{h_c}, W, \boldsymbol{A}, \boldsymbol{S})}{V(R_{hx_c}, R_{h_c}, W, \boldsymbol{S})} \tag{10}$$

where $V(R_{hx_c}, R_{h_c}, W, \boldsymbol{A}, \boldsymbol{S})$ is the number of proton pairs between the neighboring residues R_{hx_c} and R_{h_c} that have at least one assignment possibility in $\boldsymbol{A}$, and whose minimum distance among the input structures is within T_{cut}. $V(R_{hx_c}, R_{h_c}, W, \boldsymbol{S})$ is the number of such proton pairs ignoring whether or not they have an assignment possibility. The sequence window W was set to 1.

Finally, $\theta(\boldsymbol{c})$ is a bias term that favors assignments for intra-residue and sequential contacts.

Filtering by assignment score and structure

From the set of all assignment possibilities, low scoring assignments not found in any of the structures based on a 6 Å distance cutoff were removed. Low scoring assignments found in at least one structure were not filtered out. High scoring assignments not found in any of the structures were also not filtered out to allow for assignments of contacts not found in any input structure. The score cutoff was computed dynamically by first sorting the assignments in descending order by score and then selecting the cutoff where $Y\%$ of the current set of top scoring assignments were not satisfied by any of the structures. Y was chosen to be 40 %. Assignment possibilities with score less than the cutoff were filtered out unless satisfied by at least one structure.

The seed assignments from the One-to-One Seed Assignment step were also filtered similarly, except that seed assignments must be satisfied by at least one structure; otherwise, the seed assignment was converted to an ambiguous assignment. Additionally, seed assignments were filtered for isolated assignments between pairs of residues r_1, r_2 that have no other assignments for atoms between r_1 and r_2, and also no assignments for other residues nearby r_1 and r_2 based on a sequence window. The number of ambiguous assignments were also filtered to limit the number of possibilities using the concept of partial assignments similar to the idea in the program ARIA (Nilges et al. 1997). Unless stated otherwise, the same filters in the pipeline were used for all targets.

One-to-one seed assignment

A one-to-one seed assignment linking one peak to one contact was generated using a graph-based algorithm. The assignment problem was modelled after the problem of finding the maximum weighted matching on a bipartite graph (Burkard et al. 2012). The seed assignment algorithm maximizes the sum of the assignment scores subject to the constraint that each peak is assigned to at most one contact, and each contact is assigned to at most one peak. Similar restraints are found in MOTOR (Schieborr et al. 2013), but implemented differently here. Linear programming was used to solve the bipartite graph problem. The linear program is given in Eq. 1. The assignment variable $X(p, c)$ is equal to 1 if NOE peak p is assigned to contact c; otherwise it is 0. The top equation gives the sum of the scores of the assignments, and this equation is maximized over the set of all assignment variables denoted by the vector X. The second equation restricts each peak to being assigned to at most one contact, and the third equation restricts each contact to being assigned to at most one peak. Although the variables take on binary values, the restraints have the property that they form a totally unimodular matrix because the problem is formulated as the standard bipartite matching problem, so it can be solved efficiently by linear programming rather than by integer linear programming, i.e.

$$\begin{aligned} &\max_X \sum_{p,c\in A} Score(p,c) * X(p,c) \\ &\forall p, \sum_c X(p,c) \le 1 \\ &\forall c, \sum_n X(p,c) \le 1 \\ &X(p,c) \in \{0,1\} \end{aligned} \tag{11}$$

Distance calibration

After the structure-based filtering, the seed assignment was used for distance calibration. The calibration factor k in $d_p = k/I_p^6$ was determined similarly as in ARIA (Nilges et al. 1997), where $k = \sqrt{k_i k_j}$ was divided into separate

proton classes i and j depending on the proton types in the contact. Separate calibration factors were determined for backbone protons, beta protons, methyl protons, and other side-chain protons. Intra-residue and sequential contacts from the seed assignments were used for calibration, where distances with low standard deviation in the input structures were used to provide the reference distances. Similar to CANDID (Herrmann et al. 2002), the calibration factors were automatically adjusted such that at most 10 % of the distance bounds were violated by the input structures. After calibration, the S_{int} score terms were computed and then the seed assignment was recomputed.

Assignment output

The output of NOE assignment are proton–proton distance restraints, but for sampling efficiency, I-TASSER uses a reduced amino acid representation consisting of the C_α atom and a two-rotamer approximation for both C_β and the side-chain center of mass pseudoatom of the remaining side-chain heavy atoms. NMR-I-TASSER also includes the backbone nitrogen atom, which was reconstructed from the backbone C_α's by REMO (Li and Zhang 2009). The proton–proton restraints were converted to residue-based restraints by grouping proton–proton contacts that have the same residue-based contact type. The set of residue contact types $\mathbf{T}$ is given by $\mathbf{T} = (C_\alpha$–C_α, C_α–C_β, C_α–N, C_α–SC, C_β–C_β, C_β–N, C_β–SC, N–N, N–SC, SC–SC), where SC is the side-chain center of mass. Given a proton–proton contact $\boldsymbol{c}$, we shall denote its residue contact type as $t_{\boldsymbol{c}}$. The distance violation penalty function for residue contact type t between residues r_1 and r_2 is defined as

$$P(r_1, r_2, t) = \sum_{(\boldsymbol{p},\boldsymbol{c})\,|(R_{h_{\boldsymbol{c}}}, R_{h_{\boldsymbol{c}}})=(r_1,r_2),\mathrm{t_c}=t} Score(\boldsymbol{p}, \boldsymbol{c}) \qquad (12)$$

where the summation is over all peak-contact assignments where the contact is of residue contact type t and the contact contains residues r_1 and r_2. The upper bound distance UB was set to the maximum distance among the proton–proton contacts in the summation plus a correction factor for SC type contacts to account for the uncertainty of using a side-chain pseudoatom. If the distance in the structure is greater than the upper bound UB plus a 1.0 Å error tolerance, the penalty is set to $P(r_1, r_2, t)$. Otherwise, the penalty is decreased linearly until a distance of UB at which the penalty is 0. For each ambiguous assignment, the penalty and upper bound distance were set to the maximum among the assignment possibilities.

Constrained I-TASSER structure sampling

NMR-I-TASSER samples different structure conformations using a modified Monte Carlo simulation algorithm based on replica-exchange parallel hyperbolic sampling (Zhang et al. 2002). The idea of replica-exchange is to help the low-temperature simulations to cross the energy barriers that separate different minima, by swapping the conformations periodically with that found in the high-temperature simulations; and the idea of parallel hyperbolic sampling is to flatten the energy function so that energetically inaccessible regions are easier to reach. The simulations start from the fragments excised from multiple threading templates and are guided by contact restraints from both the templates and NOE assignment. The non-NOE restraints are represented by a knowledge-based energy function along with energy terms derived from statistics from the PDB. For details of the energy terms, refer to the references (Yang et al. 2015; Zhang et al. 2003).

The Monte Carlo sampling is based on fragment assembly consisting of bond-vector and fragment movements. The template fragments have variable size in order to adapt to the degree of homology between the target sequence and template structure. One of the advantages in the I-TASSER sampling is that the templates can often be drawn closer to the native by the combination of multiple threading alignments. But for hard targets that have no consensus templates, many regions need to be constructed from scratch where external restraints such as those from the NOE assignments are particularly useful to guide the sampling.

Due to the ruggedness of the energy landscape, however, the structure sampling can get trapped in incorrect conformations even when some correct restraints are provided because the incorrect template conformations may be consistent with many of the restraints (Snyder et al. 2005). To help address the issue, the structure sampling was conducted with two different sets of restraints from the templates. One set included the long-range contacts (sequence separation $\geq$6) from the high-rank templates to consider the case that the templates are correct; and the other set excluded these high-rank contacts to consider the contacts from the low-rank templates. The latter contact sets are not turned on until later in the simulations. After the clustering step, the cluster centroid structures from both runs were merged and filtered by contact overlap.

Modeling selection by decoy clustering

After structure sampling, the structure samples were clustered according to structure similarity by SPICKER (Zhang and Skolnick 2004b). In comparison to the traditional molecular dynamics simulation methods, I-TASSER does not require the sampling to converge to a single cluster. Using clustering and multiple templates can result in multiple clusters, especially in situations where there are multiple structures that are consistent with the restraints.

The cluster centroid models were filtered by contact overlap. The overlap values were converted to Z-scores by subtracting the mean and dividing by the standard deviation. Z-score cutoffs of 0, 0.5, and 1.0 were used to filter out the models in iterations 1, 2, and 3, respectively. If the top-ranked model was similar to the top-ranked model from the previous iteration, i.e. the TM-score (Zhang and Skolnick 2004a) between two generations of models is >0.9, then the top-ranked model was returned. Otherwise, the remaining cluster centroid models served as input for both NOE assignment and Monte Carlo simulations in the next iteration.

Results and discussion

NMR-I-TASSER participated in the blind assessment only for two targets in Round 2: HR8254A (PDB: 2m2e) and HR2876C (PDB: 2m5o). For other targets in the two CASD-NMR rounds, we run the pipeline retroactively by excluding all template proteins whose PDB deposition date occurs after the release date of the peak lists of the CASD-NMR target to simulate our participation for those targets. The RMSD values reported are between the backbone heavy atoms of the model (N, CA, C) and the representative conformer of the ensemble of models in the PDB file in the well-defined regions (average pairwise RMSD of the models <2 Å). For the Round 1 targets, we used the residue ranges in CASD-NMR report (Rosato et al. 2012). The TM-score (Zhang and Skolnick 2004a) and the global distance score (GDT-TS) (Zemla 2003) are also reported for the predicted models. Unlike RMSD, TM-score is less sensitive to outliers and more sensitive to the global fold topology because the metric weighs close atom pairs more than the outliers, so excluding residue ranges is not necessary. The GDT-TS score is also not sensitive to outliers because GDT only counts the percentage of residue pairs below a set of distance cutoffs (1, 2, 4, 8 Å). For brevity, each target shall be referred to by their PDB ID rather than by their target name.

Results using I-TASSER alone

It is of interest to see how a protein structure prediction program like I-TASSER performs without NMR data to determine if it can produce models useful for molecular replacement by NMR. Tables 1 and 2 give the I-TASSER results for the Round 1 and 2 targets, respectively. The data show that I-TASSER predicted the correct topology for most of the targets although the best model was not necessarily ranked first.

I-TASSER ranks the cluster centroid models according to the cluster size. For Round 1 targets, 5 out of 10 proteins have the first-ranked model <4 Å to native. If the best in the top 10 ranked models are considered, 7 out of 10 targets have a RMSD < 4 Å. For Round 2, 7 out of 10 targets have the top-ranked model <4 Å to native, and 9 out of 10 targets if the best in top 10 are considered. For two targets, 2kmm and 2l9r, I-TASSER built high-resolution structures with a RMSD < 1 Å. Figure 3 illustrates the top-ranked model for 2kmm, where is nearly a perfect match to the native. For these targets, there is little room for NMR-I-TASSER to improve, partly due to its resolution of reduced modeling. But there are several targets (e.g. 2kky, 2knr, and 2loj), where none of the top 10 models have an acceptable fold, which provides a test bench on how the NMR restraints could improve the results.

Overall, the average RMSD of the I-TASSER models in Round 1 is 5.29 Å, slightly larger than that in Round 2 (3.18 Å), partly due to the more distant template identification by LOMETS, as demonstrated by the best TM-score of the threading templates (i.e. TM-score = 0.514 for Round 1 and 0.568 for Round 2). As shown in Tables 1 and 2, the I-TASSER models have a higher TM-score in both rounds (0.564 and 0.650) than the templates, indicating the ability of I-TASSER for threading template refinements.

We note that there is one target (2kj6) in Round 1, whose PDB structure was indicated to be problematic (Zhang et al. 2014). The I-TASSER model and the NMR-I-TASSER model have only a modest similarity to the PDB structure at around 4 Å. In these models, the main helix is packed in the hydrophobic core even though NOE assignment yielded very few contacts between the helix and the core. But in the PDB structure, the helix is not packed, which was revealed later to have errors in the chemical shift assignment (Zhang et al. 2014).

NMR-I-TASSER results

As a control to the I-TASSER models, the NMR-I-TASSER modeling results are listed in the right-hand side columns in Tables 1 and 2 for the Rounds 1 and 2 targets, respectively. In general, the use of distance restraints from NOEs improved the I-TASSER modeling except in 3 cases when the I-TASSER models have a high resolution (<2 Å), which may indicate a resolution limit of using residue-based contacts. For NMR-I-TASSER, the average RMSD for the Round 1 targets was 2.14 Å compared to 5.29 Å for I-TASSER. For the Round 2 targets, the average RMSD for NMR-I-TASSER was 1.79 Å on the raw peak lists, and 1.71 Å on the refined peak lists, in control with the I-TASSER models that have an average RMSD of 3.18 Å.

Figure 4 illustrates an example from 2lci that was built on the raw peak lists. The I-TASSER model appears to have correct topology (3.65 Å), but NMR-I-TASSER was

Table 1 I-TASSER and NMR-I-TASSER results for Round 1 targets

Target	Length	I-TASSER only						NMR-I-TASSER		
		R_1 (Å)[a]	TM_1[b]	GDT_1[c]	R_{10} (Å)[d]	TM_{10}[e]	GDT_{10}[f]	R_1 (Å)[a]	TM_1[b]	GDT_1[c]
2kif	102	1.84	0.862	0.846	1.84	0.862	0.846	1.73	0.845	0.817
2kj6[g]	97	3.97	0.566	0.539	3.94	0.566	0.544	3.92	0.571	0.544
2kk1	135	4.93	0.474	0.422	2.82	0.648	0.594	1.53	0.667	0.609
2kkx	102	9.38	0.490	0.466	3.64	0.600	0.583	2.84	0.718	0.657
2kky	102	9.87	0.509	0.476	8.60	0.509	0.476	2.29	0.755	0.681
2knr	118	12.39	0.265	0.237	9.51	0.325	0.305	3.07	0.758	0.672
2kmm	73	0.80	0.813	0.843	0.77	0.823	0.843	0.99	0.785	0.815
2kpm	105	3.48	0.505	0.464	3.47	0.514	0.464	2.05	0.588	0.557
2kpt	148	4.75	0.470	0.389	4.75	0.472	0.392	1.51	0.735	0.661
2kru	63	1.50	0.684	0.754	1.44	0.684	0.754	1.44	0.671	0.726
Average	104.5	5.29	0.564	0.543	4.08	0.600	0.580	2.14	0.709	0.674

[a] R_1: Backbone heavy atom (N, CA, C) root-mean-square deviation (RMSD) between the first model (ranked by cluster size) and the representative conformer of the native model

[b] TM_1: TM-score between the first model and the representative conformer of the native model

[c] GDT_1: GDT-TS score between the first model and the representative conformer of the native model

[d] R_{10}: The minimum RMSD between the top-ten models and the representative conformer of the native model

[e] TM_{10}: The maximum TM-score between the top-ten models and the representative conformer of the native model

[f] GDT_{10}: The maximum GDT-TS score between the top-ten models and the representative conformer of the native model

[g] This PDB structure was later revealed to be problematic (Zhang et al. 2014)

Table 2 I-TASSER and NMR-I-TASSER results for Round 2 targets

Target	Length	I-TASSER Only						NMR-I-TASSER					
								Raw peak lists			Refined peak lists		
		R_1 (Å)[a]	TM_1[b]	GDT_1[c]	R_{10} (Å)[d]	TM_{10}[e]	GDT_{10}[f]	R_1 (Å)[a]	TM_1[b]	GDT_1[c]	R_1 (Å)[a]	TM_1[b]	GDT_1[c]
2l9r	69	0.54	0.750	0.801	0.54	0.782	0.826	0.90	0.710	0.765	0.99	0.690	0.732
2la6	99	1.95	0.757	0.745	1.85	0.763	0.755	1.52	0.769	0.725	1.57	0.759	0.707
2lah	160	3.08	0.790	0.719	3.00	0.794	0.722	2.25	0.788	0.664	2.18	0.790	0.700
2lci	134	3.65	0.610	0.528	3.42	0.651	0.567	1.87	0.829	0.735	1.69	0.844	0.750
2ln3	83	3.23	0.628	0.639	2.30	0.690	0.705	1.52	0.769	0.753	1.65	0.763	0.747
2loj	63	7.23	0.336	0.377	5.88	0.338	0.377	2.54	0.414	0.472	2.20	0.508	0.560
2ltm	107	4.22	0.632	0.579	3.42	0.646	0.594	1.77	0.780	0.717	1.86	0.777	0.715
2ltl	119	4.03	0.584	0.523	3.60	0.605	0.538	1.79	0.740	0.660	1.63	0.763	0.691
2m2e[g]	71	1.82	0.672	0.729	1.82	0.672	0.641	1.82	0.672	0.729	1.59	0.737	0.761
2m5o[g]	97	2.02	0.744	0.770	2.02	0.744	0.680	1.96	0.738	0.702	1.70	0.751	0.767
Average	100.2	3.18	0.650	0.641	2.79	0.668	0.640	1.79	0.721	0.692	1.71	0.738	0.713

[a] R_1: Backbone heavy atom (N, CA, C) root-mean-square deviation (RMSD) between the first model (ranked by cluster size) and the representative conformer of the native model

[b] TM_1: TM-score between the first model and the representative conformer of the native model

[c] GDT_1: GDT-TS score between the first model and the representative conformer of the native model

[d] R_{10}: The minimum RMSD between the top-ten models and the representative conformer of the native model

[e] TM_{10}: The maximum TM-score between the top-ten models and the representative conformer of the native model

[f] GDT_{10}: The maximum GDT-TS score between the top-ten models and the representative conformer of the native model

[g] Blind result from NMR-I-TASSER only. Refinement models from Xplor-NIH were submitted to CASD-NMR but not shown here

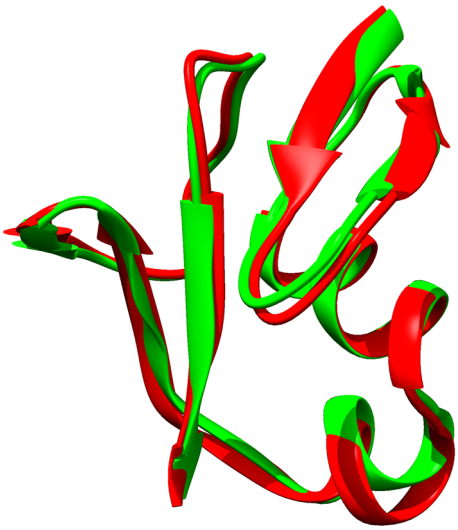

Fig. 3 Superposition of the I-TASSER model (*green*) to the native structure of 2kmm (*red*). Dynamic residues at the C-terminus have been omitted (S63-H73)

able to improve further the model (1.87 Å), especially in the beta sheet region. Although the improvements were subtle, this example shows that the coarse-grained modeling based on residue-based contacts can produce fine-grained improvements.

Fig. 4 Modeling results for 2lci. The I-TASSER model is shown *green* cartoon, the NMR-I-TASSER model from raw peak lists in *yellow*, and the native PDB representative conformer in *red*. The highlighted beta sheet region shows the improvement of the NMR-I-TASSER model over the I-TASSER-only model

The cases where I-TASSER did not succeed but NMR-I-TASSER succeeded can be placed into two groups: (a) the best I-TASSER model was not in the first (largest) cluster where NMR-I-TASSER refine both the rank and quality of the models, and (b) none of the top 10 I-TASSER cluster centroid models had the correct topology where NOE data helps NMR-I-TASSER to construct better models.

Figure 5 illustrates an example of Group (a) refinement from 2kk1. The first-rank I-TASSER model (blue) had the circled helix at the incorrect location while the 5th best-ranked model had the helix at the correct location. The first-ranked NMR-I-TASSER model (yellow) correctly placed the helix (probably by pulling the lower-rank I-TASSER models to a higher-rank), which also improved the overall quality of the best I-TASSER model, particularly on the

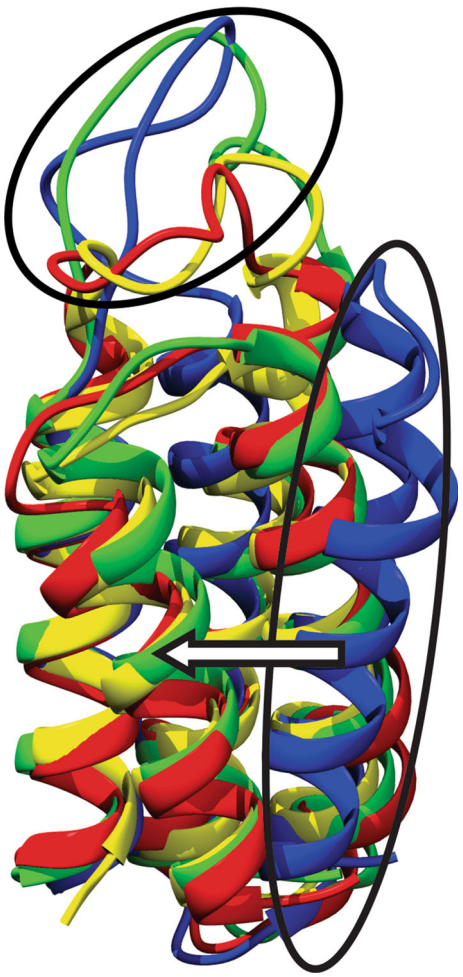

Fig. 5 Modeling results for 2kk1. The native PDB representative conformer is shown in *red* cartoon. The first-ranked I-TASSER model is in *blue*, where the circled helix (S38-L53) is at the wrong location. The 5th ranked I-TASSER model is shown in *green*, where the helix is at the correct location. The first-rank NMR-I-TASSER model shown in *yellow* improves upon the 5th ranked I-TASSER model in the circled loop region (A108-G114). The dynamic N-terminal residues M1-K36 have been removed for clarity

circled loop region in Fig. 5. The overall RMSD of the first NMR-I-TASSER model is 1.53 Å, compared to the RMSD of 2.82 Å from the best I-TASSER model.

Figures 6 and 7 present two examples of the Group (b). In Fig. 6, the first-rank I-TASSER model for 2kky had the N-terminal helix in the wrong position (Fig. 6a). NOE assignment yielded the inter-helical contacts shown as yellow lines. The final NMR-I-TASSER model has the helix relocated to the correct position. Note that although the target was classified as hard by LOMETS and all templates are incorrect on the N-terminal helix location, the NOE assignment algorithm was able to correctly assign some of the inter-helical contacts as ambiguous restraints, which eventually guided the NMR-I-TASSER simulations to draw the helix to the right location, resulting in a RMSD/TM-score improvement from 8.6 Å/0.509 to 2.29 Å/0.755 for the first NMR-I-TASSER model (Fig. 6b).

In Fig. 7a, the best I-TASSER model for 2knr is shown in green. Although the global topology was incorrect with a TM-score = 0.325, the secondary structures of the model appear generally correct but were assembled incorrectly. The model in Fig. 7b shows the first-ranked NMR-I-TASSER model after the first iteration (magenta), and Fig. 7c shows the final-iteration model (yellow). The final model improved upon the first iteration model at the C-terminal tail region as highlighted by the upper circle in Fig. 7c. Both NMR-I-TASSER models have the topology generally correct except for the lower loop region. In both iterations, the upper bound distances between the left and right parts of the loops (highlighted by the lower circle in Fig. 7c) were too large, so they are farther apart in the model compared to in the native structure. This problem of distance calibration highlights the current NOE assignment which needs to be further addressed.

In these two examples, the use of two different sets of template restraints—one set that included the long-range template contacts and one set without—improved the modeling because the different structure samplings increased the chance of finding better models. For 2kky, the set that included the long-range template contacts had the best overall model at 3.29 Å (TM-score = 0.668) from the 5th largest cluster after the first iteration, and the model was ranked the highest based on the contact overlap criterion. The best model from the excluded set was 3.89 Å (TM-score = 0.595), but both models were automatically selected for the next iteration based on contact overlap along with various other models, which resulted in a much improved final model of RMSD/TM-score = 2.29 Å/0.755 in the final iteration. Contrarily, for 2knr the simulations with the long-range template contacts excluded had the better model (RMSD = 4.02 Å) than that without (5.04 Å). Contact overlap ranked the 4.02 Å model higher, but both models were automatically selected, which increases the chance of model selection for the next iteration.

Blind test results

The protocol used for the blind tests is slightly different from that described in Fig. 1 because the blind tests were performed in 2013. The main difference is that the NOE assignment-based filtering of templates was not implemented at that time, and the input models for NOE assignment came from running I-TASSER alone rather than from threading. In addition, all-atom refinement was performed with Xplor-NIH (Schwieters et al. 2003). Nevertheless, the results in Tables 1 and 2 show only the NMR-I-TASSER result before the all-atom refinement step.

For 2m2e, LOMETS classified the threading result as Easy with all models having 3 alpha helices similar to the Myb-like DNA-binding domain (Fig. 8). The first-ranked I-TASSER model without NOE restraints had an RMSD of 1.82 Å. The raw peak lists were noisier than many of the

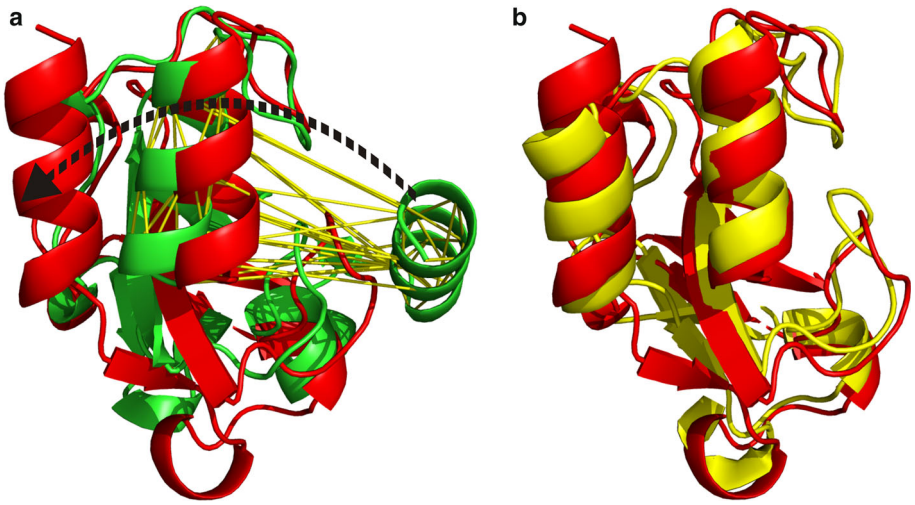

Fig. 6 Modeling results for 2kky. The native PDB representative conformer is shown in *red* cartoon. **a** The *green* cartoon represents the best I-TASSER model superposed on the native, with lines highlighting the contacts from the NOE assignment step in the first iteration. Only the contacts between the two helices are shown for clarity. **b** The final NMR-I-TASSER model shown in *yellow* has the helix (S1-C12) at the correct location

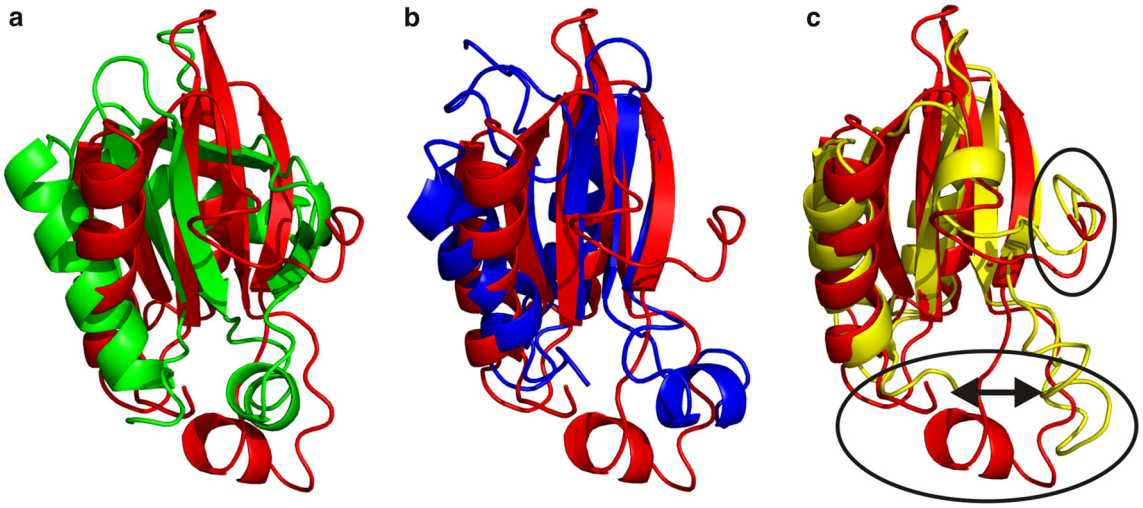

Fig. 7 Modeling results for 2knr. The native PDB representative conformer is shown in *red* cartoon. **a** The *green* cartoon represents the best I-TASSER model overlaid on the native. **b** The magenta cartoon shows the top-ranked NMR-I-TASSER model selected by contact overlap after the first iteration. **c** The *yellow* cartoon give the final NMR-I-TASSER model. The *upper circle* highlights the main improvement in the C-terminal region starting from E105 to the C-terminus. In the *lower circled* region, there are restraints from NOEs between the two contacting segments (R2-L3 with R65-E67), but the upper bound distances used are too large, so the segments are farther apart in the model than in the native structure

other targets in this round. For example, the number of peaks in 2m2e is greater than that for the target 2lah, despite 2lah contains more than twice as many residues. NOE assignment from the raw peak lists yielded some obvious unphysical long-range contacts (the bending of the C-terminal helix by more than 90°), which caused NMR-I-TASSER to bend the helix unless the restraints are removed. Therefore, we chose to refine directly from the

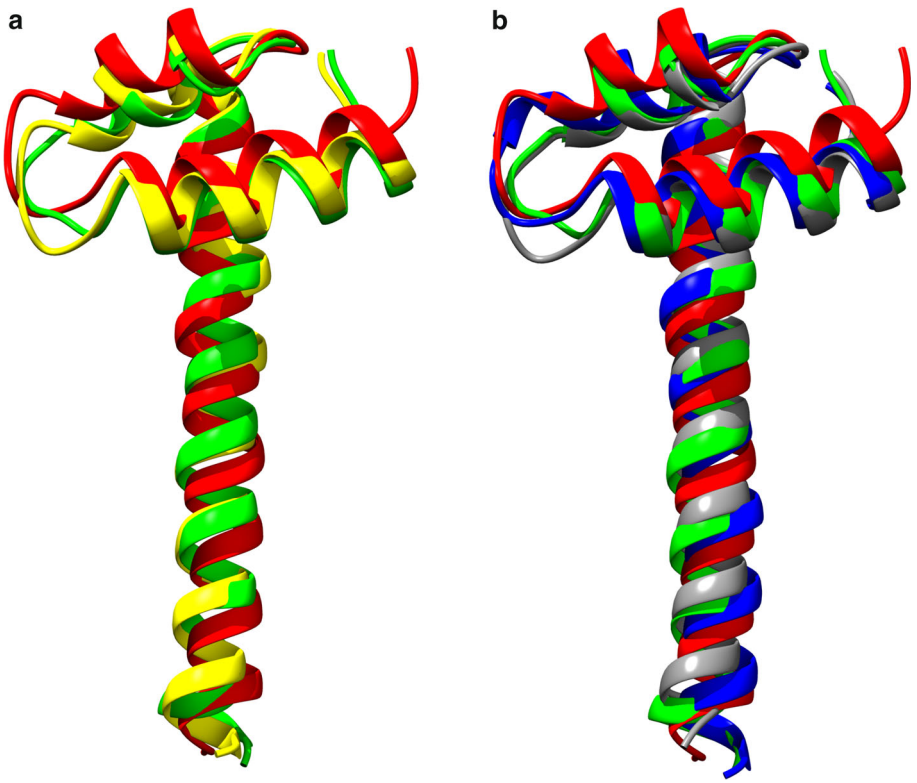

Fig. 8 Blind testing results of target 2m2e. All models are shown as cartoons where the native PDB representative conformer is shown in *red* and the I-TASSER only model is in *green*. **a** From raw peak lists, final model after all-atom refinement (*yellow*). This model was refined directly from the I-TASSER only model. **b** From refined peak lists, NMR-I-TASSER model (*grey*) and final model after all-atom refinement (*blue*)

I-TASSER-only model. For refinement, we used the I-TASSER model to filter out the unphysical long-range contacts. The lowest energy refined model from the raw peak lists improved to an RMSD of 1.54 Å (Fig. 8a). This example illustrates that the I-TASSER model can be used to effectively filter out noise in the peak lists. From the refined peak lists, the RMSD of the NMR-I-TASSER model was 1.59 Å and after refinement it was reduced to 1.45 Å (Fig. 8b).

Similar to 2m2e, the I-TASSER assembly simulation constructed a good initial model for 2m5o with an RMSD of 2.02 Å for the next step NMR-I-TASSER refinement (Fig. 9). By combing the raw peak list data, NMR-I-TASSER produced the first model of a RMSD 1.96 Å (Fig. 9a), which was further improved to 1.70 Å when using the refined peak lists (Fig. 9b). After all-atom refinement with the raw peak list data, the RMSD of the first model was reduced to 1.36 Å (Fig. 9a); and it became 1.00 Å (if the first 3 residues from the N-terminal tail are excluded) with the refined peak lists (Fig. 9b). For this target, dynamic residues as predicted by the chemical shift assignment-based TALOS+ program (Shen et al. 2009) were removed prior to the structure calculation to improve sampling efficiency. However, for simplicity of comparison to PDB structures, the dynamic residues were not removed in non-blind tests.

To conclude this section, we briefly summarize the all-atom refinement protocol. To obtain the NOE restraints, our NOE assignment algorithm was used. Three iterations of refinement were performed by first generating 50 models, taking the top 10 by energy, running the NOE assignment algorithm with these 10 models as input, and then filtering the contacts not satisfied by at least 1, 2, or 5 models (based on a 6 Å distance cutoff) depending on whether it is the first, second, or third iteration respectively. The final model selected was the lowest energy model according to Xplor-NIH.

Comparison of modeling with raw and refined peak lists

Round 2 saw the introduction of raw peak lists in addition to the refined ones. To evaluate the difference, we compared the NOE assignment accuracy for each peak list using the top 10 models from running I-TASSER alone as the input structures (see Table 3). In general, the accuracy of NOE assignments was better with the refined peak lists. For the long-range ($\geq$6 sequence separation) seed assignments, the average accuracy was 85.0 and 92.8 % for the raw and refined peak lists respectively. For the ambiguous assignments, if we define a correct assignment as that with at least one contact possibility in the ambiguous assignment being correct, the average accuracy was 85.8 and 96.9 % for the raw and refined peak lists, respectively.

For some targets, there was a large difference in the NOE assignment accuracies, so the quality of the raw peak lists appears to vary significantly across the targets. For example, the accuracy of the seed assignment of 2la6 did not vary much: 96.4 % for raw peak lists and 96.7 % for refined. However, the difference between the raw and refined peak lists for 2lci, 2loj, and 2m2e was large: from 56.7 to 74.0 % for 2lci, 73.8 to 85.5 % for 2loj, and 71.9 to 95.3 % for 2m2e. Coincidentally, the lowest assignment accuracies for the raw peak lists were for these 3 targets. For 2m2e, we observed that it had many more peaks compared to the other targets even though the protein has only 71 residues, which indicates that more aggressive filtering by the NOE assignment algorithm should be used to better filter the seed assignment for this target. For 2lci, although the raw peak list seed assignment accuracy was low, the ambiguous assignment accuracy was high (83.8 %), so a model <2 Å was produced.

Despite the differences in NOE assignment accuracy, with respect to the final structure models, there does not appear to be a significant difference in quality between the models from raw or refined peak lists, which have an

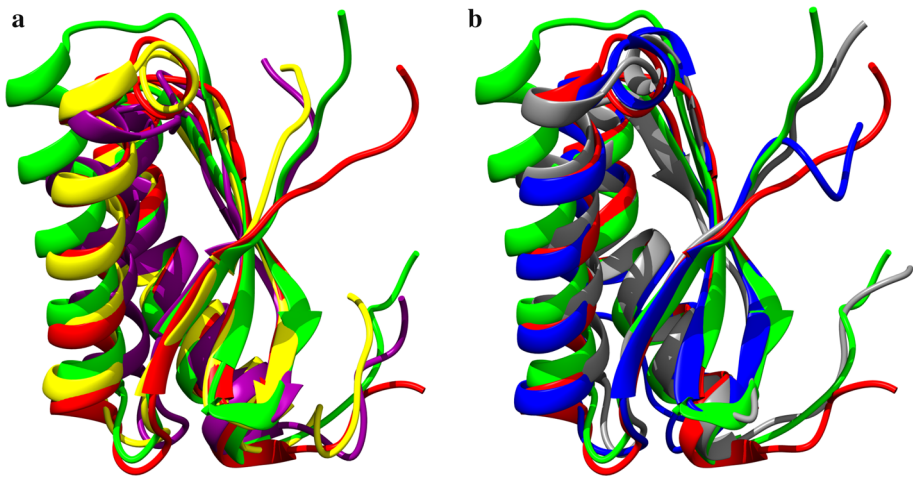

Fig. 9 Blind testing results of target 2m5o. All models are shown as cartoons where the native PDB representative conformer is shown in *red* and the I-TASSER only model is in *green*. **a** From raw peak lists, NMR-I-TASSER model (*purple*) and final model after all-atom refinement (*yellow*). **b** From refined peak lists, NMR-I-TASSER model (*grey*) and final model after all-atom refinement (*blue*)

Table 3 NOE seed assignment results from raw and refined peak lists for Round 2 targets based on I-TASSER model input

Targets	Accuracy[a]		Count[b]		Count/residue[c]	
	Raw	Refined	Raw	Refined	Raw	Refined
2l9r	1.00	1.00	125	152	1.8	2.2
2la6	0.96	0.97	607	659	6.1	6.7
2lah	0.94	0.96	476	718	3.0	4.5
2lci	0.57	0.74	323	358	2.4	2.7
2ln3	0.91	0.92	275	412	3.3	5.0
2loj	0.74	0.86	103	76	1.6	1.2
2ltm	0.88	0.98	632	452	5.9	4.2
2ltl	0.87	0.96	219	202	1.8	1.7
2m2e	0.72	0.95	430	85	6.1	1.2
2m5o	0.92	0.95	411	455	4.2	4.7
Average	0.85	0.93	360.1	356.9	3.6	3.4

[a] Fraction of long-range seed assignments (sequence separation ≥ 6) with C$\alpha < 6$ Å in at least 50 % of the conformers in the native PDB structure

[b] Total number of long-range seed assignments

[c] Total number of long-range seed assignments divided by the number of amino acids

average RMSD of 1.79 and 1.71 Å, respectively. This indicates that the structure-sampling algorithm is relatively robust to the additional assignment errors from using raw peak lists. Furthermore, the number of restraints derived from raw peak lists is not significantly more than that from refined peak lists. The fraction of all contacts with sequence separation ≥ 6 in the native structure that are found in the seed assignment is 0.20 and 0.19 for raw and refined peak lists, respectively.

Conclusion

NMR-I-TASSER is a hierarchical template-based method that takes a molecular replacement-like approach to the NMR structure determination, where the initial NOE assignment step uses structure information from threading templates. The approach takes advantage of multiple templates and produces multiple clusters of models using a reduced residue-based conformational search. Even though the conformational search is course-grained, 8 out of 10 Round-2 targets had high-resolution models (<2 Å), and this was obtained without performing all-atom refinement. For two targets from 2kmm and 2l9r, the I-TASSER models were both <1.0 Å, which was even more accurate than the NMR-I-TASSER model due to errors in the NOE assignment. Although limited in resolution due to the coarse-grained modeling, we have shown that NMR-I-TASSER can produce accurate starting models in the conformational search. Therefore we believe the pipeline can play a complementary role to the existing all-atom NMR structure determination methods. For example, for proteins larger than 10 kDa, a combination of traditional NMR methods with all-atom refinement was shown to be more CPU efficient and just as effective as restrained assembly simulations (Mao et al. 2014). It is known that for large proteins, traditional NMR methods may not converge, so the combination of NMR-I-TASSER with the traditional all-atom refinement methods will be an interesting direction to explore, in particular for efficient and effective structure determination of large proteins.

NMR-I-TASSER is potentially useful in cases when there is sparse data, where multiple structure models may be consistent with the data. In these cases, it is important to quickly generate many different structure topologies, especially if the structure sampling needs to sample away from incorrect template conformations. The problem then becomes more of a model selection problem, which requires a good function to discriminate between models that fit and those that do not fit the NMR data well. For ranking models, we have thus far ignored important chemical shift information. Adding a score based on chemical shift prediction from structure, such as from ShiftX2 (Han et al. 2011), and combining it with different model quality scores, such as the DP-score (Huang et al. 2005) and MQA score (Kosciolek and Jones 2014) should prove helpful for model selection, in addition to the inherent I-TASSER force field.

Acknowledgments This work was supported in part by National Science Foundation Career Award (DBI1027394); National Institute of General Medical Sciences (GM083107 and GM084222); National Natural Science Foundation of China (30700162); Fundamental Research Funds for the Central Universities of China (2014TS138); and China Postdoctoral Science Foundation (2014M552043). Molecular graphics were performed with the UCSF Chimera package (Pettersen et al. 2004).

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bowie JU, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253:164–170

Burkard R, Dell'Amico M, Martello S (2012) Assignment problems. Other titles in applied mathematics. Society for Industrial and Applied Mathematics, Philadelphia

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763

Han B, Liu Y, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 50:43–57

Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new

software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209–227
Huang YJ, Powers R, Montelione GT (2005) Protein NMR recall, precision, and F-measure scores (RPF Scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc 127:1665–1674
Kosciolek T, Jones DT (2014) *De Novo* structure prediction of globular proteins aided by sequence variation-derived contacts. PLoS One 9:e92197
Li Y, Zhang Y (2009) REMO: a new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks. Proteins 76:665–676
Li W, Zhang Y, Skolnick J (2004) Application of sparse NMR restraints to large-scale protein structure prediction. Biophys J 87:1241–1248
Li Y, Roy A, Zhang Y (2009) HAAD: a quick algorithm for accurate prediction of hydrogen atoms in protein structures. PLoS One 4:e6701
Mao B, Tejero R, Baker D, Montelione GT (2014) Protein NMR structures refined with rosetta have higher accuracy relative to corresponding X-ray crystal structures. J Am Chem Soc 136:1893–1906
Meiler J, Baker D (2003) Rapid protein fold determination using unassigned NMR data. Proc Natl Acad Sci USA 100:15404–15409
Montelione GT (2014) NMR experimental data. In: 11th community wide experiment on the critical assessment of techniques for protein structure prediction, Riviera Maya, Mexico
Nilges M, O'Donoghue SI (1998) Ambiguous NOEs and automated NOE assignment. Prog Nucl Magn Reson Spectrosc 32:107–139
Nilges M, Macias MJ, O'Donoghue SI, Oschkinat H (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin. J Mol Biol 269:408–422
Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. doi:10.1002/jcc.20084
Rosato A et al (2012) Blind testing of routine, fully automated determination of protein structures from NMR data. Structure 20:227–236
Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738. doi:10.1038/nprot.2010.5
Schieborr U et al (2013) MOTOR: model assisted software for NMR structure determination. Proteins Struct Funct Bioinform 81(11):2007–2022. doi:10.1002/prot.24361
Schwieters CD, Kuszewski JJ, Tjandra N, Marius Clore G (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160:65–73
Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223
Snyder DA, Bhattacharya A, Huang YJ, Montelione GT (2005) Assessing precision and accuracy of protein structures derived from NMR data. Proteins 59:655–661
Wishart DS, Arndt D, Berjanskii M, Tang P, Zhou J, Lin G (2008) CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. Nucleic Acids Res 36:W496–W502
Wu ST, Zhang Y (2007) LOMETS: a local meta-threading-server for protein structure prediction. Nucl Acids Res 35:3375–3382
Wu S, Zhang Y (2008) MUSTER: improving protein sequence profile-profile alignments by using multiple sources of structure information. Proteins 72:547–556
Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein structure and function prediction. Nat Methods 12:7–8
Zemla A (2003) LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Res 31:3370–3374. doi:10.1093/nar/gkg571
Zhang Y, Skolnick J (2004a) Scoring function for automated assessment of protein structure template quality. Proteins 57:702–710
Zhang Y, Skolnick J (2004b) SPICKER: a clustering approach to identify near-native protein folds. J Comput Chem 25:865–871
Zhang Y, Kihara D, Skolnick J (2002) Local energy landscape flattening: parallel hyperbolic Monte Carlo sampling of protein folding. Proteins 48:192–201
Zhang Y, Kolinski A, Skolnick J (2003) TOUCHSTONE II: a new approach to Ab initio protein structure prediction. Biophys J 85:1145–1164
Zhang J, Liang Y, Zhang Y (2011) Atomic-level protein structure refinement using fragment guided molecular dynamics conformation sampling. Structure 19:1784–1795
Zhang Z, Porter J, Tripsianes K, Lange OF (2014) Robust and highly accurate automatic NOESY assignment and structure determination with Rosetta. J Biomol NMR 59:135–145