

# Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex

Diego F. Gauto<sup>1</sup>, Leandro F. Estrozi<sup>1</sup>, Charles D. Schwieters<sup>2</sup>, Gregory Effantin<sup>1</sup>, Pavel Macek<sup>1,6</sup>, Remy Sounier<sup>1,7</sup>, Astrid C. Sivertsen<sup>1</sup>, Elena Schmidt<sup>3</sup>, Rime Kerfah<sup>1,6</sup>, Guillaume Mas<sup>1,8</sup>, Jacques-Philippe Colletier<sup>1</sup>, Peter Güntert<sup>3,4,5</sup>, Adrien Favier<sup>1,9</sup>, Guy Schoehn<sup>1</sup>, Jerome Boisbouvier<sup>1,9</sup>, & Paul Schanda<sup>1,9</sup>

<sup>1</sup>*Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), 71, avenue des martyrs, F-38044 Grenoble, France.*

<sup>2</sup>*Center for Information Technology, National Institutes of Health, Bethesda, USA.*

<sup>3</sup>*Institute of Biophysical Chemistry, Goethe University Frankfurt am Main, Frankfurt am Main, Germany.*

<sup>4</sup>*Laboratory of Physical Chemistry, ETH Zürich, Zürich, Switzerland.*

<sup>5</sup>*Graduate School of Science, Tokyo Metropolitan University, Tokyo, Japan.*

<sup>6</sup>*Present address: NMR-Bio, 5 place Robert Schumann, F-38025 Grenoble, France.*

<sup>7</sup>*Present address: Institut de Génomique Fonctionnelle, CNRS UMR-5203 INSERM U1191, University of Montpellier, F-34000 Montpellier, France*

<sup>8</sup>*Present address: Biozentrum University of Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland.*

<sup>9</sup>*correspondence to be addressed to: [adrien.favier@ibs.fr](mailto:adrien.favier@ibs.fr), [jerome.boisbouvier@ibs.fr](mailto:jerome.boisbouvier@ibs.fr), [paul.schanda@ibs.fr](mailto:paul.schanda@ibs.fr)*

**Atomic-resolution structure determination is the key requirement for understanding protein function. Cryo-EM and NMR spectroscopy both provide structural information, but currently cryo-EM does not routinely give access to atomic-level structural data, and, generally, NMR structure determination is restricted to small (<30 kDa) proteins. We introduce an integrated structure determination approach that simultaneously uses NMR and EM data to overcome the limits of each of these methods. The approach enabled determination of the high-resolution structure of the 468 kDa large dodecameric aminopeptidase TET2 to a precision and accuracy below 1 Å by combining secondary-structure information obtained from near-complete magic-angle-spinning NMR assignments of the 39 kDa-large subunits, distance restraints from backbone amides and specifically labelled methyl groups, and a 4.1 Å resolution EM map. The resulting structure exceeds current standards of NMR and EM structure determination in terms of molecular weight and precision. Importantly, the approach is successful even in cases where only medium-resolution (up to 8 Å) cryo-EM data are available, thus paving avenues for the structure determination of challenging biological assemblies.**

For decades, the weight of atomic-resolution structure elucidation, a requirement for understanding biomolecular mechanisms in detail, has been almost exclusively borne by X-ray crystallography, as highlighted by the fact that about 90 % of all entries in the Protein Data Bank (PDB) are crystal structures. Notwithstanding this success of crystallography, many supra-molecular edifices, self-assembling systems, membrane proteins and proteins with extended dynamic domains are difficult to crystallize, or the crystals do not diffract to high resolution. Single-particle cryo electron microscopy (cryo-EM) and nuclear magnetic resonance spectroscopy (NMR) are not bound to obtaining well-ordered crystals, and are applicable even in the presence of significant motions. Enabled by decisive instrumental and methodological progress, cryo-EM has recently made a leap in resolution [\[12\]](#), which opened avenues for determining protein structures at atomic detail. Nonetheless, *de novo* atomic-resolution structure determination from cryo-EM data is currently not the rule. A survey of the electron density maps deposited in the EM data base (EMDB) over the last two years reveals an average resolution in the range of 6.5-8 Å. In 2016, a particularly productive year for single-particle cryo-EM (961 entries in the EMDB), only 20% of the deposited maps had a resolution of 3.5 Å or better, thus being well suited for atomic-level *de novo* structure determination. Extending EM structure determination to a wider range of biological objects may require further resolution increase and/or combination of EM with other data.

NMR spectroscopy probes the structure of proteins on a length scale that is very different from that seen by EM. Rather than probing the electronic potential of the molecule, providing a more or less well-defined molecular envelope, NMR detects the immediate vicinity of hundreds to thousands of individual atomic nuclei across the protein. Their resonance frequency is exquisitely

sensitive to the local electronic environment and, thus, to the local structure. Furthermore, dedicated NMR experiments identify which atoms are in the vicinity ( $< 5\text{-}10\text{ \AA}$ ) of a given atomic nucleus, thus allowing to assemble structural elements into a full three-dimensional structure. The precision and accuracy of NMR-derived structures strongly depend on the availability of a large number of such distance restraints across the entire molecule. When NMR structure determination fails, it is generally the paucity of meaningful distance restraints – often as a consequence of resonance overlap or low detection sensitivity – which hampers the arrangement of local structural elements to a well-defined tertiary fold. In solution-state NMR, structure determination is particularly challenging for proteins larger than 50 kDa, as the slow molecular tumbling leads to rapid nuclear relaxation and thereby low detection sensitivity, which results in difficulties obtaining information about local conformation and distances. Consequently, besides three cases<sup>3-5</sup> the PDB is, to our knowledge, devoid of *de novo* NMR structures beyond 50 kDa (subunit size). The introduction of deuteration and specific methyl labeling with  $^{13}\text{CH}_3$  groups has greatly expanded the molecular weight range in which solution-NMR signals can be detected to ca. 1 MDa.<sup>6-8</sup> Yet because this approach is limited to methyl groups, no information about the backbone conformation can be obtained, which excludes any *de novo* structure determination.

Magic-angle spinning (MAS) solid-state NMR, overcomes the inherent size limitations that its solution-state counterpart faces by replacing the slow stochastic molecular tumbling in solution – which is at the origin of the rapid NMR signal loss – by an immobilization of the protein, and spinning of the sample at a constant "magic" angle.<sup>9</sup> Recent advances in MAS NMR sample preparation, including optimized isotope-labeling schemes, hardware allowing fast sample spin-



ning (>50 kHz), and optimized pulse sequences have enabled the structure determination of small to medium-sized proteins forming crystals, oligomeric assemblies, or embedded in membranes, generally with monomer sizes below 20 kDa,<sup>[10]</sup> with an exceptional case at 27 kDa.<sup>[11]</sup> Sensitivity limitations and the increased spectral complexity and signal overlap in large proteins have so far been a bottleneck for the extraction of sufficient useful restraints for structure determination on larger proteins.

While EM and NMR spectroscopy probe protein structure on different length scales, these complementary views may be combined. Recent work has explored the joint use of NMR and EM data to determine protein structures.<sup>[12-14]</sup> The proteins studied are small (< 100 residues) and comprise only 2 or 3 secondary structure elements, which made it possible to either first determine a global fold from NMR and refine this model with the aid of the EM map<sup>[12]</sup>, or to build a first model into the electron density map, and use NMR data to refine these manually constructed EM models<sup>[13][14]</sup>. For larger proteins, the situation is significantly more complex: the former approach, i.e., determining a rough fold from NMR and then refining with EM data, is challenging because lack of NMR restraints—often the case for large proteins—precludes obtaining a converged structural model which then could be used for EM-based refinement. The latter method suffers from the fact that *de novo* building of the protein sequence into an electron density, i.e., unambiguously identifying the residues and being able to follow the chain throughout the electron density, becomes increasingly complex with protein size, and very strongly depends on the resolution of the EM data.

Here, we present an approach that overcomes these challenges and combines NMR and EM information in a joint and lightly-supervised manner. Briefly, the key ingredients for the combined analysis from these two techniques are (i) the identification of structural features (such as  $\alpha$ -helices) in 3D space from EM maps, (ii) the NMR-based identification of these structural elements along the sequence, in particular the residue-wise assignment of secondary structures, (iii) the unambiguous assignment of these sequence stretches to 3D structural features detected by EM, guided by NMR-derived distance restraints, and (iv) the joint refinement of the protein structure against NMR data and an electron-density map.

We show the feasibility of this approach to large protein assemblies with the example of the dodecameric  $12 \times 39$  kDa (468 kDa) large aminopeptidase enzyme TET2 from *Pyrococcus horikoshii* which forms a large hollow lumen encapsulating twelve peptidase active sites.<sup>[15][16]</sup> We use EM maps with resolutions of 4.1, 6 and 8 Å as well as secondary structure and distance information from MAS NMR and solution-NMR. Despite its large subunit size of 353 residues, larger than any protein assigned to date by MAS NMR, we achieved a near-complete NMR resonance assignment in a reliable manner, which allows the residue-wise determination of the backbone conformation, as well as hundreds of experimental distance restraints involving the protein backbone and side chain methyl groups, but by themselves these data are insufficient for structure determination. An automated approach that jointly exploits the electron density map, NMR secondary structure and distance restraints is presented, that allowed us to obtain the structure of TET2 at a backbone root-mean-square-deviation, RMSD, of 0.7 Å to the crystal structure.

Importantly, we show that the NMR/EM approach provides near-atomic-resolution structures even if only moderate-resolution EM data (up to 8 Å resolution) are available, an often-encountered situation that prohibits structure determination from EM data alone. Our structure provides insight into regions that had not been visible in previous crystal structures of TET2, in particular a conserved loop region in the catalytic chamber of this enzyme machinery.

## Results

### High-resolution NMR and cryo EM provide structural information on different length scales.

Any investigation of structure, interactions or dynamics by NMR-based methods is bound to the availability of well-resolved spectra and the ability to assign the observed peaks to individual atoms along the protein sequence. Figure [1A](#) and B show MAS NMR spectra of the protein backbone amides and solution-NMR of Ile and Val methyl groups in TET2, respectively. Despite the large number of residues per TET2 subunit, we obtained well-resolved spectra and high sensitivity. In solution, methyl spectra can be observed that are very similar to the corresponding MAS NMR spectra obtained. This shows that the sample states are structurally comparable, which allows combining the data from solution-NMR and MAS NMR in a straightforward manner (see Figure [S1](#)). To obtain atom-specific assignments of these spectra we collected a series of three- and four-dimensional  $^{13}\text{C}$ -detected and  $^1\text{H}$ -detected MAS NMR assignment spectra, which correlate the frequencies of neighboring atoms, thereby allowing to connect them along the protein backbone and out into the side chains (Figure [S2A, B](#)). Additional spectra recorded on three different samples with amino-acid-type specific  $^{13}\text{C}$ -labeling (either only LKP (Leu, Lys, Pro), GYFR or

ILV residues) provided simplified spectra, which served as convenient starting points for manual assignment (see Figure S2C-F, Tables S2 to S4 and Methods for details). Together, these spectra were sufficient to obtain by manual analysis a near-complete assignments of TET2, ca. 85% of the backbone and 70% of the side chain heavy atoms (Figure S3). Residues for which no assignments were obtained are restricted to N- and C-termini and loops, a finding that can be attributed to the lower efficiency of the MAS NMR experiments for such dynamic parts. The assignment of 69 Ile, Leu, Val methyl groups was achieved through a combination of a mutagenesis-based strategy in solution, reported previously<sup>17,18</sup>, and the <sup>13</sup>C frequencies obtained from MAS NMR assignment experiments. Even though TET2 is larger than any protein assigned to date to a similar extent by MAS NMR, the approach is sufficiently robust to obtain the near-complete assignment even in a fully automatic manner with the program FLYA<sup>19</sup> (Figure S4), showing that reliable assignments can be obtained without time-consuming manual spectra analysis. The assigned chemical shifts allow the determination of  $\phi$  and  $\psi$  backbone dihedral angles, and thus the secondary structure from a database approach, TALOS-N<sup>20</sup>. NMR-detected secondary structure elements are displayed in Figure 1D.

In order to gain tertiary structure information, we measured short inter-atomic contacts through solution-NMR 3D nuclear Overhauser effect spectroscopy (NOESY) spectra and a 3D MAS NMR radio-frequency driven recoupling (RFDR) experiment (Figure S5). The solution-NMR data identify spatial proximity between Ile <sup>$\delta^1$</sup>  and Val<sup>proS</sup> methyl groups; a second NOESY spectrum, recorded on a sample in which different subunits were labeled in different methyl groups, furthermore allowed filtering out inter-subunit methyl-methyl contacts (Figure S5E-G).

The MAS NMR experiment provides distances between backbone amide sites and between amides and methyls of Ile, Val and Leu side chains, as well as contacts between such methyls in a single 3D MAS NMR experiment.  $^{13}\text{C}$ - $^{13}\text{C}$  distances were additionally measured in the selectively LKP-labeled sample. Figure 1E and Table 1 show all distance restraints identified by NMR, including 471 spectrally unambiguous ones, i.e., of which the cross-peak frequencies pointed unambiguously to one atom pair and 45 spectrally ambiguous ones, i.e., which could not be assigned to a single atom pair (but which could be unambiguously assigned along the structure calculation procedure, see below).

Attempts to determine the TET2 subunit structure from these NMR data using a standard structure-determination protocol based on restrained MD, CYANA<sup>21</sup>, failed in achieving convergence to the correct structure (backbone RMSD to the mean of the ten lowest-energy structures and to the crystal structure above 10 Å, see Figure S6). In these NMR-only structure calculations, many local structural elements, e.g., the local packing of two  $\alpha$ -helices or  $\beta$ -strands, are formed, but the positions of these elements relative to each other in the tertiary fold remained poorly defined, thus calling for additional data.

Figure 1C shows the cryo-EM map of TET2 at 4.1 Å resolution, calculated from 27,130 single particles with the software RELION (see Methods section for details and Figure S7). We have attempted to build the structure of TET2 from this electron density map using *phenix.model\_to\_map*<sup>22</sup>, in the crystallographic software suite Phenix.<sup>23</sup> The program was able to trace the map, with better success when the symmetry of the particle was calculated from the cryo-EM map rather than in-

ferred from the biological unit derived from the X-ray structure. However, the map quality was too low for phenix to assign the sequence and thus build a model (Figure S8). While it likely would have been possible to solve the structure at 4.1 Å by careful manual building, or by using other tools, the failure of this workhorse of crystallography model building illustrates the complexity of building a *de novo* model even at a resolution of 4.1 Å, which currently is among the high-resolution range for cryo-EM. This finding also highlights that, in all likelihood, succeeding with model building rapidly evaporates with decreasing resolution.

Despite the inability to trace the backbone chain in the density, a number of features, in particular  $\alpha$ -helices, are readily recognized in the electron density, even at lower resolution (up to 8 Å) and in a fully automatic unsupervised manner<sup>24</sup>. Figures 1F, G show the  $12 \times 7$   $\alpha$ -helices that could be automatically detected from the electron density map of TET2 truncated to 8 Å resolution, and Figure S9A shows additionally detected  $\beta$ -sheet regions. NMR and EM thus provided structural information about TET2 on very different length scales, but neither of the two techniques by themselves allowed to determine the structure at the atomic scale, owing to the lack of long-range distance information in NMR, and the inability to build a robust model from the electron density.

**An automated approach for integrated EM/NMR structure calculation.** We developed an approach for the combined use of EM and NMR data, keeping in mind that it should be applicable to cases where only medium-resolution EM maps (between 6 and 10 Å) would be available. Its central idea is the fact that certain structural elements can be identified even in such medium-

resolution EM maps and by unsupervised automatic algorithms. The arrangement of these structural elements in 3D space, i.e., the knowledge of the positions of even a small subset of residues relative to each other, potentially brings very useful long-range structural information, and it is exactly such long-range information that would be of enormous benefit in NMR-based structure calculations.

Placing sequence stretches into the density can be viewed as a combinatorial assignment problem, in which sequence elements with known properties (e.g.,  $\alpha$ -helices,  $\beta$ -strands or large side chains that can be recognized in EM maps) are assigned to electron-density features with the matching shape. The most straightforward case is the one of  $\alpha$ -helices, given the ease of detecting the  $\alpha$ -helical electron-density features automatically, and assigning  $\alpha$ -helical structures to sequence stretches from NMR data (Figure 1F,G and 1D, respectively).

Briefly, our approach, outlined in Figure 2, consists of three steps. In the first step, after detection of the  $\alpha$ -helices, the various possibilities of placing  $\alpha$ -helical sequence stretches into  $\alpha$ -helix density features are systematically and combinatorially tested, performing with each of these helix-to-density assignments a restrained-MD type structure calculation with CYANA. This type of calculation follows standard procedures adopted in any NMR-based structure calculation, but the key difference is that in addition to NMR-derived distance restraints the  $\alpha$ -helices are placed at precisely defined positions relative to each other. Subsequently, the correct helix-to-density assignment is identified through objective criteria that compare the resulting structural models to experimental NMR and EM observables. With the correct helix-to-density assignment at hand,

steps 2 and 3 of our approach refine this model by iterative addition of restraints from the EM map as well as from NMR distance restraints that could be unambiguously assigned based on the initial structural models (see below).

An important consideration for the combinatorial procedure of step 1 is that the number of possible helix-to-density assignment combinations increases exponentially with the number of  $\alpha$ -helices. Considering that the eight  $\alpha$ -helices found in the sequence by NMR can be placed into the automatically detected helix features in the map (seven per subunit), including two orientations in each helix, there are  $8! \times 2^7 \approx 5 \cdot 10^7$  possibilities (see Table S1 and discussion therein). Even more combinations are possible when considering that it is not straightforward to unambiguously identify which helical densities belong to the same subunit. Performing structural calculations with each of these assignments would be computationally very expensive.

A computationally manageable number of assignments can be achieved by taking into account some simple experimental observables. First, we used a cluster analysis to identify  $\alpha$ -helical densities which are closest in space as belonging to the same asymmetric unit, and ascribed these five helical densities to the same subunit (see Figure 1G and Figure S9). Furthermore, we explicitly disregarded the polarity when assigning helical stretches to densities, by enforcing that only the central residue of a given  $\alpha$ -helical sequence stretch is fixed to the center of an  $\alpha$ -helical density, rather than the entire helix. Lastly, we discarded all those assignments for which the lengths of the  $\alpha$ -helices along the sequence did not match the length of the helical density (see Table S1). With these assumptions and simplifications, only forty possible assignments remain and we per-



formed forty CYANA structure calculations with these helix-to-density assignments (five residues restrained at fixed positions relative to each other), along with the backbone dihedral-angle and the spectrally unambiguous NMR distance restraints.

Identification of the correct helix-to-density assignment was based on two criteria, (i) the CYANA target function, which reflects the agreement of the structure with the distance and dihedral-angle restraints, and (ii) the overlap of the five  $\alpha$ -helical stretches with the electron density. Incorrect helix-to-density assignments generally result in bent or twisted  $\alpha$ -helices, and thus low overlap with the EM map and violation of NMR distances. Figure 2B displays these two scores for the twenty CYANA runs with the lowest CYANA score. A funnel-shaped distribution is observed for the various structures. Figure 2C displays two of these structures, in which either one or five  $\alpha$ -helices were assigned correctly. The structure with the correct helix-to-density assignment scores best among all structures. While it has the topology approximately correct, its precision and accuracy are low, calling for further refinement steps.

The refinement stage of the protocol, an automated iterative procedure denoted here as Step 2 (Figure 2D), consists of iterative cycles of (i) a flexible fitting of the initial structure into the 8 Å EM density with iMODFIT<sup>25</sup>, and (ii) CYANA structure calculations. In each iteration, EM and NMR restraints were added: on the one hand, sequence stretches consistently located inside the 8 Å electron density were restrained to their positions, in the same manner as the helix-centers were initially fixed relative to each other (see Figure S10); on the other hand, the initial structural models allowed the unambiguous assignment of more NMR distance restraints, by excluding as-

segment possibilities corresponding to physically unrealistic long-range distances (see Methods). This iterative procedure was repeated until convergence of the bundle-RMSD (Figure 2E). In Step 3, a real-space structure refinement was applied in the XPLOR-NIH software<sup>26,27</sup>, using the full electron density map, first at 8 Å, then at 6 Å and lastly at 4.1 Å resolution, and all NMR restraints, calculating the full dodecamer. Figure 3 shows the evolution of the structure from the initial model to the final structure, which was obtained with high precision (bundle backbone RMSD over the 20 lowest-energy structures of 0.29 Å for one subunit).

The refined dodecameric structure (Figure 4) is in excellent agreement with the crystal structure. The dodecameric assembly forms a tetrahedral structure with ca. 12 Å wide openings, one on each of the four tetrahedron faces, providing access to the ca. 50 Å wide catalytic lumen. Our structure features a long, hitherto unobserved loop region in the catalytic chamber, comprising residues 119-138, shown in green in Figure 4C. Interestingly, even though the crystal structure has significantly higher resolution than the EM map (1.75 Å, 16 vs. 4.1 Å), this structural feature is much more clearly visible in the latter (Figure 4). Moreover, while this loop region appears relatively well defined at cryo-EM conditions, it is not observable at room temperature by MAS NMR, and we were unable to assign the resonances of the backbone of this loop (Figure S3E). These findings may suggest that the loop is dynamically disordered at room temperature, leading to low signal intensity in MAS NMR experiments, and to non interpretable density in crystal structures determined at cryogenic temperatures.

An important consideration is how our approach performs with lower-resolution EM data.

At 4.1 Å resolution, our EM map is better than the average of the maps deposited in recent years (average in 2018: 6.6 Å). We repeated our structure determination approach using maps truncated to either 6 Å or 8 Å in the refinement step. (Note that in Steps 1 and 2, our approach is anyhow based on a 8 Å map only.) Figures 4F and 4G show the refined structures with these two lower-resolution maps, which still have a decent accuracy, with backbone RMSD values to the crystal structure of 1.8 and 2.6 Å, respectively (calculated over  $C\alpha$ ). We additionally explored a refinement with *phenix.real\_space\_refine*<sup>28</sup> to refine these structures furthermore, and obtained an even slightly higher structure accuracy (see Table S6 and Methods). These results highlight that the combined EM/NMR approach is able to provide near-atomic-resolution structures even in cases where the resolution is worse than the present-day average values, and for which *de novo* structure determination is currently far out of reach using EM methods only.

## Discussion

The recent "resolution revolution" in cryo-EM is progressively changing the way how the structures of biomolecular assemblies are determined, but currently *de novo* structure determination remains reserved to a minority of EM data sets. Even with the comparably favorable 4.1 Å resolution of our TET2 EM map we were unable to trace the polypeptide chain *de novo* using currently available tools (see Figure S8). The integration of complementary data from different techniques holds great promise for obtaining atomic-resolution structures even when any single method provides too sparse data. The present work is among the few examples of such an integration of EM and NMR data. Compared to these previous structures determined from NMR/EM data<sup>12-14</sup>,

TET2 has a more than four times larger monomer size, being larger than any protein for which near-complete MAS NMR resonance assignments have been reported so far.

Our approach uses features that can be identified in a straightforward manner even in medium (8-10 Å) resolution EM maps, and in an automatic unsupervised manner, most commonly  $\alpha$ -helices. Although not exploited here, also some  $\beta$ -sheet regions and bulky side chains could be readily identified in the EM maps (cf. Figure S9A) and exploited similarly as the  $\alpha$ -helices used here. The inclusion of only very few such real-space restraints—in our case only five points along the sequence—turns out to be extremely powerful in driving the structure calculation towards the correct fold. Specific isotope-labeling schemes, in particular deuteration and methyl-labeling, were crucial as they allowed the unambiguous assignment of hundreds of NMR-distance restraints between residues distant in primary sequence, thereby enabling to use the NMR-score to identify the correct helix-to-density assignment, which is the first step to guide the structure towards the correct topology. A remarkable result of this study is the finding that even an EM map at 8 Å resolution allowed us to determine the fold accurately, whereas the *de novo* modeling even with a 4.1 Å resolution map failed.

Data from other sources can be readily included into this integrated approach. The distance restraints obtained here from solution-state and MAS NMR may be complemented or replaced by distance information obtained from co-evolution data<sup>[29-31]</sup>, cross-linking experiments,<sup>[32]</sup> Förster resonance energy transfer,<sup>[33]</sup> DEER electron paramagnetic resonance<sup>[34-36]</sup>, or NMR-detected paramagnetic relaxation enhancement (PRE) data.<sup>[37]</sup> The secondary structure information obtained here

from NMR chemical-shift assignments may be complemented with prediction algorithms in case of missing NMR assignments. Integrating such diverse data in structure determination will likely play an important role in structural studies of challenging biological assemblies forming oligomeric structures, such as capsids, tubes/needles/fibers, i.e., structures that generally are very difficult to study by crystallography.

## Methods

### 1 Sample preparation

**Expression and isotope-labeling of TET2.** TET2 from *Pyrococcus horikoshii* (UniProt entry O59196) was obtained by overexpression in *E. coli* BL21(DE3) cells with suitable isotope-labeled M9 minima, supplemented with  $^{15}\text{NH}_4\text{Cl}$  in all cases as sole nitrogen source. Depending on the desired labeling scheme (outlined below), suitably labeled D-glucose (uniformly  $^{13}\text{C}_6$ -labeled or 1,2,3,4,5,6- $^2\text{H}_7$ ,  $^{13}\text{C}_6$ -labeled or 1,2,3,4,5,6,6- $^2\text{H}_7$ -labeled or unlabeled) was used as carbon source, and for the case of samples with amino-acid specific labeling, isotope-labeled amino acids or ketoacid precursors were added before induction with isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG, 1 mM in final culture).

For MAS NMR, six different samples were used in this study with different labeling. Sample 1: u- $^{13}\text{C},^{15}\text{N}$ . Sample 2: u- $^{15}\text{N}$ -Gly,Tyr,Phe,Arg- $^{13}\text{C}$ . Sample 3: u- $^{15}\text{N}$ -Ile,Leu,Val- $^{13}\text{C}$ . Sample 4: u- $^{15}\text{N}$ -Leu,Lys,Pro- $^{13}\text{C}$ . Sample 5: u- $^2\text{H},^{13}\text{C},^{15}\text{N}$ . Sample 6: u- $^2\text{H},^{15}\text{N}$ , Ile $^{\delta 1}$ ,Leu $^{\delta 2}$ ,Val $^{\gamma 2}$ - $^{13}\text{CHD}_2$ , i.e.,  $^{13}\text{CHD}_2$  methyl groups at Ile- $\delta^1$ , Val- $\gamma^{\text{proS}}$  and Leu- $\delta^{\text{proS}}$  positions in an otherwise perdeuterated background (the pro-S methyl is the one denoted as  $\gamma 2$  in valine and as  $\delta 2$  in leucine).

For sample 1,  $^{13}\text{C}_6$ -labeled glucose was used. For samples 2-4, unlabeled D-glucose was used. For sample 3  $^{13}\text{C}_4$ -2-ketobutyrate (60 mg/L of culture) was added 20 minutes prior to induction and  $^{13}\text{C}_5$ -acetolactate (300 mg/L) was added 1 h prior to induction. The resulting Ile residues have  $^{13}\text{C}$  labels at the  $\text{C}^\alpha$ , CO,  $\text{C}^{\gamma 1}$  and  $\text{C}^{\delta 1}$  sites; the leucines bear  $^{13}\text{C}$  at the  $\beta$ ,  $\gamma$  and  $\delta$  sites; the

valines are uniformly  $^{13}\text{C}$  labeled<sup>[38]</sup>. For samples 2 and 4, the u- $^{13}\text{C},^{15}\text{N}$ -labeled amino acids were added 1 h prior to induction.<sup>[39]</sup> For sample 5,  $\text{D}_2\text{O}$ -based M9 medium was used, with  $^2\text{H}_7,^{13}\text{C}_6$ -labeled glucose. For sample 6, Ile $^{\delta 1}$ , Leu $^{\delta 2}$ , Val $^{\gamma 2}$  sites were labeled by addition of 2-ketobutyrate (60 mg/L; 20 min prior to induction) and acetolactate (200 mg/L; 1 h prior to induction), with  $^{13}\text{CHD}_2$ -groups at the relevant positions<sup>[38,40]</sup>. For the production of the deuterated, methyl-labeled samples, cells were progressively adapted, in three stages, over 24 h, to M9/ $\text{D}_2\text{O}$  media containing 1 g/L  $^{15}\text{ND}_4\text{Cl}$  and 2 g/L  $^2\text{H}_7$  glucose (Isotec). In the final culture, the bacteria were grown at  $37^\circ\text{C}$  in M9 media prepared with 99.85%  $\text{D}_2\text{O}$  (Eurisotop).

Samples 1–4 were used for resonance assignment with  $^{13}\text{C}$ -detected MAS NMR experiments (see Table S2). Sample 5 was used for  $^1\text{H}$ -detected MAS NMR assignment experiments, listed in Table S4. Sample 6 was used to collect  $^1\text{H}$ - $^1\text{H}$  distance restraints between methyl and amide groups by RFDR MAS NMR (see Table S5). Sample 4 was furthermore used to collect  $^{13}\text{C}$ - $^{13}\text{C}$  distance restraints.

For solution-state NMR NOESY experiments, a uniformly deuterated sample with  $^{13}\text{CH}_3$  groups at Ile- $\delta^1$  and Val- $\gamma^{\text{proS}}$  sites was used (sample 7). We refer to this labeling as u- $^{13}\text{CH}_3$ , u- $^{15}\text{N}$ , Ile $^{\delta 1}$ , Val $^{\gamma 2}$ - $^{13}\text{CH}_3$ ). This sample was used for collecting methyl-NOESY spectra, listed in Table S5. Three additional samples were used for the identification of intermolecular contacts, labeled either with u- $^{13}\text{CH}_3$ , u- $^{15}\text{N}$ , Ile $^{\delta 1}$ , Thr $^{\gamma 2}$ , Val $^{\text{proS}}$ - $^{13}\text{CH}_3$  (sample 8), or u- $^{13}\text{CH}_3$ , u- $^{15}\text{N}$ , Met $^{\epsilon}$ , Ala $^{\beta}$ , Val $^{\text{proR}}$ - $^{13}\text{CH}_3$  (sample 9), and a sample in which these two types of subunits were re-assembled (sample 10), as described in Figure S5E-G. In the present work, the detection of inter-subunit contacts, however, turned out to be not critical, and these three samples and the according NOESY data sets

were used only to exclude eight inter-subunit contacts.

**Purification of TET2.** Cells were harvested by centrifugation and broken in a microfluidizer (15 kpsi, 3 passes) in 50 mM Tris-HCl/150 mM NaCl buffer. The total protein extract from disrupted cells was heated to 85°C for 15 minutes and the lysates cleared by centrifugation (17000 g, 1 h, 4°C). The supernatant was loaded onto a 6-ml Resource Q column (GE Healthcare) equilibrated in 100 mM NaCl, 20 mM Tris-HCl, pH 7.5. After washing with 3 column volumes, the protein was eluted with a linear salt gradient (0.1 - 0.35 M NaCl in 20 mM Tris-HCl, pH 7.5), followed by a size exclusion chromatography step (HiLoad 16/60 Superdex 200, GE Healthcare). Sample purity was verified by SDS-PAGE, and protein-containing fractions were pooled and concentrated.

**NMR samples.** Samples used for the solution-state NMR NOESY experiment were at a protein subunit concentration of 1 mM (i.e., a concentration of dodecameric particles of 83  $\mu$ M) in a D<sub>2</sub>O-based buffer (20 mM Tris, 20 mM NaCl, pH 7.4) in a 5 mm Shigemi tube.

Samples for MAS NMR were prepared by concentrating TET2 to 10 mg/mL in H<sub>2</sub>O-based buffer (20 mM Tris, 20 mM NaCl, pH 7.4) and then adding 2-Methyl-2,4-pentanediol (MPD), a commonly used crystallization agent, in a 1:1 (v/v) ratio. A white precipitate (possibly of nanocrystalline nature) appeared immediately after mixing of the solutions. The samples were filled into either 1.3 mm (Bruker), 1.6 mm (Agilent), 3.2 mm Agilent or 3.2 mm (Bruker) rotors, using an ultracentrifuge filling device adapted for a Beckman SW32 ultracentrifuge. Typically, samples were filled for ca. 1 h at ca. 50000 x g. Rotors were glued with two-component epoxy glue to avoid loss of solvent.



**Sample for cryo-electron microscopy.** A sample of u-[ $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-labeled TET2 was used for cryo-electron microscopy. The isotope labeling, although unnecessary for cryo-EM, was useful to record ssNMR spectra with the same sample, in parallel to cryo-EM as a sample quality control. Three and a half microliters of sample were applied to 2:1 glow discharged Quantifoil holey carbon grids (Quantifoil Micro Tools GmbH, Germany) and the grids were frozen in liquid ethane with a Vitrobot Mark IV (FEI, the Netherlands).

## 2 NMR spectroscopy

All solution-NMR experiments were performed at a temperature of 50°C, either with a Varian INOVA 800 MHz spectrometer (for the H-H-C NOESY) or a 950 MHz Bruker Avance 3 spectrometer, both equipped with cryogenically cooled probes. MAS NMR experiments were performed at an effective sample temperature of ca. 28°C, measured from the bulk water resonance and the resonance of MPD at 4.1 ppm. MAS NMR assignment experiments were recorded either on a Agilent 600 MHz VNMRs spectrometer equipped with a triple-resonance 3.2 mm probe (for  $^{13}\text{C}$ -detected experiments) or a 1.6 mm Fast-MAS probe (for all reported  $^1\text{H}$ -detected experiments at a MAS frequency below 40 kHz), or a Bruker 1000 MHz Avance spectrometer equipped with a 3.2 mm probe (only one 3D experiment, see Table S2). All additional  $^1\text{H}$ -detected experiments at MAS frequencies >50 kHz were recorded with a Bruker 600 MHz Avance 3 spectrometer equipped with a 1.3 mm MAS probe tuned to  $^1\text{H}$ ,  $^2\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  frequencies. Acquisition parameters of all experiments are reported in Tables S2 to S5.

**NMR resonance assignment.** Assignment of NMR resonances by MAS NMR was performed with a suite of 3D and 4D experiments with either  $^{13}\text{C}$  detection or  $^1\text{H}$  detection, listed in Tables S2, S3 and S4.

Inspection of the spectra and manual peak picking was performed with the CCPNMR software<sup>41</sup>. Near-complete resonance assignments were achieved by manual analysis. Hereby, the two 4D spectra, CONCACB and CANCOCX, were crucial to unambiguously identify neighboring spin systems. They share three frequencies (CA, N, CO), and the combination of the two spectra thus allows unambiguously connecting five or six frequencies. Comparing patterns in these spectra, as well as in 3D NCACX, NCOCX and NCACB spectra allowed the sequential connection of such 5-tuples of frequencies. Assignments of the side chains were obtained from the three-dimensional NCACX, NCOCX and CCC experiments. The assignment of all reported heavy atoms was achieved from  $^{13}\text{C}$ -detected experiments only. The additional  $^1\text{H}$ -detected experiments were used to assign the amide- $^1\text{H}$  frequency.

Following the manual assignment we investigated the ability to obtain fully automatic assignments. Peaks in all spectra listed in Tables S2-S4 were picked manually (numbers of picked peaks are listed in Tables S2 to S4), and used as input for the automated assignment algorithm FLYA<sup>19</sup>, implemented in CYANA version 3.97<sup>21</sup>. The output of the automatic assignment procedure is shown in Figure S4.

This suite of experiments allowed the assignment of the heavy atoms ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) and amide  $^1\text{H}$ . The additional assignment of methyl groups was achieved by a mutagenesis-based solution-

NMR assignment approach, reported previously for Ile <sup>17</sup> and Val <sup>18</sup>. These assignments were confirmed and extended to Leu by the <sup>13</sup>C assignments obtained by MAS NMR.

In cases where several methyl groups have the same <sup>13</sup>C frequency, this information was insufficient by itself. Thus, we furthermore used the information from the RFDR MAS NMR experiment, in particular cross-peaks from the methyl group to the amide backbone site. All resonance assignments are reported in the BioMagResBank (BMRB ID 27211).

**NMR distance restraints measurement.** Table <sup>S5</sup> summarizes all acquisition parameters of solution-state and MAS NMR experiments used for determining distance restraints. Two types of solution-state NMR NOESY spectra were collected. In a first approach, a u-[<sup>2</sup>H, <sup>15</sup>N], Ile-[<sup>13</sup>CH<sub>3</sub>]<sup>δ1</sup>, Val-[<sup>13</sup>CH<sub>3</sub>]<sup>pro-S</sup> labeled sample was used to collect NOESY distance restraints with a 3D H-H-C edited experiment. <sup>18</sup> In order to exclude inter-subunit contacts, which might induce errors in single subunit structure calculations, we additionally measured two 3D H-C-C NOESY experiments: one on a sample in which different subunits were labeled differently (sample 10; see Figure <sup>S5E</sup>) and two samples where all subunits were labeled identically (sample 8). Eight inter-subunit contacts detected in this experiment were excluded from the calculation.

MAS NMR distance restraints were obtained through a 3D <sup>1</sup>H-<sup>15</sup>N/<sup>13</sup>C-<sup>15</sup>N/<sup>13</sup>C time-shared RFDR experiment and a 2D <sup>13</sup>C-<sup>13</sup>C DARR experiment (see Table <sup>S5</sup> and Figure <sup>S5</sup>).

### 3 Cryo-electron microscopy

The sample was observed with a FEI Polara at 300 kV. Movies were recorded manually on a K2 summit direct detector (Gatan Inc., USA) in super resolution counting mode at a nominal magnification of 20,000 (0.97 Å/pixel at the camera level) for a total exposure time of 4 s and 100 ms per frame resulting in movies of 40 frames with a total dose of  $\sim 40 \text{ e}^-/\text{Å}^2$ . 90 movies were manually recorded with Digital Micrograph (Gatan Inc., USA). Movies were first motion corrected with Digital Micrograph (Gatan Inc., USA) resulting in micrographs like the one on Figure S7. 1643 particles were picked semi-automatically with boxer from the EMAN suite<sup>[42]</sup> and 2D classified in 16 classes in RELION 1.4<sup>[43]</sup>. The best-defined 2D class averages were used to autopick particles in all the micrographs with RELION resulting in a total of 30407 particles. CTF estimation with CTFFIND3<sup>[44]</sup>, particle extraction in boxes of  $200 \times 200$  pixels and preprocessing were done in RELION as well. The data set was first cleaned by 2D classification leaving 27130 particles. From the best 2D classes, a low resolution *ab initio* 3D model was generated using the RICO server<sup>[45]</sup> imposing tetrahedral symmetry. The latter was then used as a starting model to refine all particle orientations in RELION (with tetrahedral symmetry imposed). Further 3D classifications were attempted but did not result in any improvement of the structure. The final 3D reconstruction has a resolution of 4.1 Å at FSC = 0.143.

### 4 Structure calculation

The protocol described in this work employed standard programs routinely used in NMR structure determination and EM fitting, namely CYANA (version 3.97)<sup>[21]</sup>, XPLOR-NIH (version 2.44.8),<sup>[26]</sup>

iModFit (version 1.44),<sup>[25]</sup> helixhunter2<sup>[24,46]</sup>, Phenix as well as in-house written scripts in python language (available from the authors upon request).

**Step 1a:  $\alpha$ -helix detection and assignment.** The automatic detection of the  $\alpha$ -helix and  $\beta$ -sheet regions from the cryo-EM map was done with the 8 Å resolution map using helixhunter2<sup>[24,46]</sup> and *gorgon*.<sup>[47]</sup> *helixhunter1*, was able to detect only the five longer alpha helices, while both *helixhunter2* and *gorgon* added two shorter  $\alpha$ -helices and the latter also detected  $\beta$ -sheet regions that are shown in Figure S9A. We used only the five longest  $\alpha$ -helices (and their eleven symmetry mates across the dodecamer), denoted A-E in Figure 1G and Table S1, for generating structure restraints used in Step 1 of the protocol.

While not strictly necessary in this protocol, for reasons of computational efficiency it is helpful to be able to identify which  $\alpha$ -helices belong to a given TET2 monomer, as it allows reducing the number of helix-to-density assignments if only helical densities within one subunit need to be considered. We clustered these  $12 \times 5$  helical densities into groups corresponding to individual subunits by an automatic approach based on the "density-based spatial clustering of application with noise (DBSCAN)" algorithm that was implemented in the Sklearn python package (<https://pypi.org/project/sklearn/#history>), and the outcome is shown in Figure S9. In essence, the underlying assumption is that the  $\alpha$ -helices that are closest to each other belong to the same subunit. For the case of the five helices we considered in TET2, this assumption turns out to be correct. Note, however, that our approach does not rely on the validity of the assumption that the closest helices belong to one subunit; any other combinations, including (erroneous) assignments

of one chain to densities belonging to different subunits can equally be considered in the CYANA calculations. We have tested a number of such erroneous helix-to-density assignments, and all these calculations yielded very high CYANA scores and could be discarded on this basis (data not shown). Thus, erroneous assignment of one chain to different subunits is detected through lack of convergence along the protocol.

Table [S1](#) shows the helix-to-density assignment possibilities which we have taken into account in Step 1, based on the lengths of the  $\alpha$ -helices. In addition to considering the helix lengths, all the physically meaningless assignments in which the same  $\alpha$ -helix would be placed at two or more locations at the same time are filtered out. The forty possible assignments, listed in Table [S1](#) were used to generate lists of residue numbers (the centers of the NMR-detected  $\alpha$ -helices) assigned to coordinates in space (the centers of the five  $\alpha$ -helical densities). These inter-helix distance restraints were used in the subsequent CYANA calculations.

**Step 1b: CYANA structure calculation and ranking assignments.** Forty CYANA calculations (version 3.97)<sup>[21](#)</sup> were performed, using as restraints the unambiguous distance restraints, backbone dihedral-angle restraints and the restraints of helix-centers to helical-density centers. In all CYANA calculations performed in this study, only one chain (353 residues) was used, and the dodecamer was built from twelve chains only in Step 3. Because CYANA works in dihedral-angle space in a protein-internal frame, rather than in real space, the implementation of the real-space helix restraints required translating the absolute-space information into relative-space information. In practice, distance restraints were applied between all atoms that were to be fixed in real space, e.g.,

between the five helix-centers in Step 1. For practical purposes, this method results in fixing the relative orientation and distance of these points, but leaves translational and rotational freedom to the whole protein, which is inconsequential for the resulting structure.

For each helix-to-density assignment, one-thousand conformers were computed using the standard simulating annealing protocol in CYANA with 4000 torsion-angle dynamics steps per conformer. Typical computation times for generating these 1000 conformations was ca. 8.3 minutes on an Intel 8-core desktop computer. Finally, the 10 conformers with the lowest final target energy function were retained, and their energy was used as one criterion for ranking of the assignments. This CYANA score is plotted as one of the axes in Figure 2B. Note that out of the forty CYANA runs, twenty were stopped due to divergent target function energy, showing that these helix-to-density assignments were incompatible with the other NMR data. The twenty remaining calculations are shown in Figure 2B.

The second criterion for selecting the correct assignment was the goodness of the fit between the electron density map and the structure, evaluated only for the five helices considered in a given helix-to-density assignment. Basically, the idea is to evaluate whether all residues of the helix reside within the experimental electron density; incorrect helix-to-density assignment would lead to a distorted structure and therefore the helices would not fit in the electron density.

To implement this criterion in practice, electron-density maps were generated in silico from the lowest-energy CYANA model, using the backbone N, C $\alpha$ , C and O atoms of each of the residues in the five helices which were used in the given helix-to-density assignment (as identified

by TALOS-N), using the module *molmap* in UCSF Chimera.<sup>[48]</sup> We compared these in silico maps to an experimental map, which comprised only the five  $\alpha$ -helices considered (colored and labeled A-E in Figure 1G, map truncated to 8 Å resolution). The module *measure correlation* in UCSF chimera was used to compute the correlation between the in silico maps (of each residue) and this experimental "helix-density" map. This correlation coefficient, obtained residue by residue, reflects the goodness of fit to the map. Examples are shown in Figure S10. As a global measure, we computed the percentage of residues that had a good overlap ( $> 0.7$ , in which 1.0 and 0.0 means good and bad overlap between the maps, respectively). This percentage is shown in Figure 2B as vertical axis.

**Step 2: Iterative refinement.** In this step, we performed iterative rounds of flexible fitting using a common tool used in EM (iModFit<sup>[25]</sup>) and an NMR-type structure calculation with CYANA. For the latter we used the information about the match to the EM map, similarly as the five helix centers were restrained in Step 1. This procedure is described as follows.

The structure resulting from Step 1 was taken as a seed to perform a normal-mode based flexible fitting into the density map with the software iModFit v1.44.<sup>[25]</sup> After this procedure, we identified those protein regions which resided well within the experimental electron density, in order to be able to fix those residues in the following CYANA calculations. This identification how well the residues correspond to experimental electron density was done in a similar manner as the goodness-of-fit criterion described in the preceding paragraph, using UCSF chimera modules *molmap* (allowing to generate density maps from a PDB file) and *measure correlation* (allowing the computation



with an experimental map), as follows. We created an in silico map for each residue (backbone N, C $\alpha$ , C, O only) and computed its correlation to the experimental EM map, truncated to 8 Å resolution, using the full map (rather than only the helices used above). Figure S10 shows this overlap score for each residue, as well as graphical examples. Residues with a good overlap score to the map (arbitrarily set as  $> 0.7$ ), were restrained in the following CYANA structure calculation. To fix the atoms in real space using CYANA, even though CYANA operates in internal dihedral-angle space, the same procedure was applied as in Step 1, i.e., the atoms were constrained through distance restraints to all those other atoms that were also fixed in real space (using a tolerance of  $\pm 0.5$  Å). *Notethatinpracticeitisofnoimportancewhereinabsolutespacethestructureisplaced, solongastheE*  
*atomicdistanceswasgeneratedusinganin-housewrittenpythonscript, andwasusedinthesamemannerasth*

NMR distance restraints with ambiguous assignments, i.e., resulting from cross-peaks with frequencies that could be assigned to more than one atom-pair, can be disambiguated with the help of the intermediate structure obtained after the flexible fitting. Briefly, for cases with multiple possible assignments (based on NMR chemical-shift positions) we measured the corresponding distances in the intermediate structures for all those atom pairs that were possible assignment candidates based on their frequency. In cases where out of the possible candidates only one atom pair had short a ( $< 8$  Å) distance, this assignment was retained in the subsequent CYANA calculation. Note that one may also use the restraints as ambiguous restraints in CYANA, but we found better convergence with this approach.

Furthermore, we also considered that some of the NOESY and RFDR distance restraints may be due to inter-subunit contacts. Introducing these restraints in a calculation that only considers

one chain would necessarily lead to distance violations. To identify such cases (of which we had only 8 altogether), 3D HMQC-NOESY-HMQC spectra were recorded on a sample in which differently labeled subunit types were mixed in a 1:1 ratio and which allows recognizing inter-subunit cross-peaks (see Figure S5G-I) and compared to a uniformly labeled sample (see Figure S5E-G for details). Cross-peaks identified in this way as stemming from inter-subunit contacts were excluded from further analysis.

CYANA calculations in Step 2 used this growing list of restraints from the match to the map as well as the increasing list of unambiguous NMR distances along with the NMR dihedral-angle restraints. Only one subunit was used in these calculations; the explicit dodecamer was considered only in Step 3.

**Step 3: Final real-space refinement.** In the final refinement step, all dihedral-angle and distance restraints from NMR as well as the full EM map (in real space), at a resolution of either 4.1, 6 Å (Figure 4F) or 8 Å (Figure 4G), were used in a joint calculation of the full dodecamer using XPLOR-NIH<sup>26</sup> version 2.44.8. Xplor-NIH's strict symmetry facility<sup>49</sup> was used to generate subunit coordinates with tetrahedral symmetry from those of a protomer subunit. Initially, protomer coordinates from Step 2 were moved as a rigid body to fit the full construct into the EM map, also allowing overall center of mass motion so that the dodecamer could be centered on the map. This starting structure was refined, allowing all protomer degrees of freedom under all restraints using the standard purely repulsive nonbonded energy term. In each refinement calculation 100 structures were calculated differing in random velocities given at the beginning of molecular dynamics

simulated annealing.

The spatial restraining effects of the EM map were introduced by performing calculations against progressively higher resolution maps: the 8 Å map, followed by the 6 Å map, and finally the highest-resolution (4.1 Å) map. The fit structure was used as initial coordinates for the first calculation, and the lowest energy calculated structure from each step was utilized as the initial structure for the subsequent calculation. Finally, the outcome of the calculation using the 4.1 Å map was refined in implicit solvent using the effective energy function for Xplor-NIH (EEF<sub>x</sub>)<sup>50</sup>. For generating the structures shown in Figure 4F and G, this gradual refinement process was stopped before adding the higher-resolution maps (i.e., before adding the 6 or 4.1 Å resolution maps, respectively), in order to evaluate the quality of the structures resolution from lower-resolution EM data.

In a final stage, to locally refine the structure of the Zn<sub>2</sub> active site, we introduced distance restraints to the zinc ions. The identity of these chelating side chains can be readily predicted from the sequence.<sup>51</sup> At this stage of the protocol, the chelating residues (H68, D182, E213, E235, H323) were already close in space to each other, allowing the inclusion of explicit restraints between these atoms to model the zinc site using generous distances based on those found in crystal structures. Also in this region, restraints were added from residue 323 to residues 92 and 93 based on a subsequent reanalysis of unassigned RFDR cross peaks. The final XPLOR-NIH calculation was performed in implicit solvent with the 4.1 Å resolution map, all NMR restraints and the Zn-site restraints. For the calculation with the lower-resolution maps (6 Å, 8 Å), this refinement in implicit solvent has been omitted. The structure deposited as 6f3k in the PDB corresponds to the result of

this refinement.

In addition, we also investigated whether the result can be further improved with phenix, a commonly used program in crystallography. The lowest-energy structures from the XPLOR-NIH runs, employing EM data at different resolution, were refined against the cryo-EM maps using *phenix.real\_space\_refine*<sup>28</sup>. We additionally used as restraints all distance measurements and secondary structure information from NMR. Different refinement schemes were tested, which all included five iterative cycles of rigid-body fit of individual chains and minimization of atomic positions, but also simulated annealing or morphing or local grid search. The results of these refinements are summarized in Table S6. We found simulated annealing to be superior to the two other approaches, as well as to not using it. Models obtained by performing Step 3 at 6 and 8 Å, respectively, were refined using the same approach, i.e. by five iterative cycles of real-space refinement including rigid-body fit of individual chains, minimization of atomic positions, and simulated annealing. Refinements were carried out both at 6 or 8 Å, and at 4.1 Å, to allow a fair comparison of the low resolution models with that determined by performing Step 3 at 4.1 Å. Correlation coefficients between the cryo-EM map and the model were calculated using *phenix.real\_space\_refine*, and correspond to the CCmask described by Jiang & Brunger<sup>52</sup> based on map values inside a mask calculated around the macromolecule. Root-mean-square deviations between the various models and the biological unit derived from the X-ray structure (1y0r) were calculated using the *super* function in PyMOL.

**De novo structure building from the EM map with Phenix** We attempted to rebuild the TET2 structure *de novo*, by use of the 4.1 Å cryo-EM map only, using *phenix.model\_to\_map*<sup>22</sup>, in the Phenix suite of crystallographic software<sup>23</sup>. The program was able to trace the map, with better success when the symmetry of the particle was calculated from the cryo-EM map rather than inferred from the biological unit derived from the X-ray structure. However, the map quality was too low for phenix to assign the sequence and thus build a model. We tested a variety of building schemes, the results of some of which are shown in Figure S8).

**Data deposition** The NMR assignment has been deposited in the BioMagResBank (entry 27211). The 4.1 Å EM map has been deposited in the EMDB (entry 4179). The atomic model resulting from the three steps described herein, using in the final step the 4.1 Å EM map, has been deposited in the Protein Data Bank (PDBid: 6f3k).

**Acknowledgements** This work was supported by the European Research Council (StG-2012-311318-ProtDyn2Function to P. S. and CoG-2010-260887 to J. B.) and used the platforms (NMR, EM, isotope-labeling) of the Grenoble Instruct-ERIC Center (ISBG; UMS 3518 CNRS- CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The electron microscopy facility is supported by the Rhone-Alpes Region, the Fondation Recherche Medicale (FRM), the fonds FEDER and the GIS-Infrastructures en Biologie Sante et Agronomie (IBISA). Financial support from the TGIR-RMN-THC Fr3050 CNRS for conducting the research at the 1 GHz spectrometer at the CRMN Lyon is gratefully acknowledged. J.-P.C. acknowledges support by the Agence Nationale de la Recherche (ANR-17-CE11-0018-01).

**Competing Interests** The authors declare that they have no competing financial interests.

**Correspondence** Correspondence and requests for materials should be addressed to J. B.

(jerome.boisbouvier@ibs.fr), A. F. (adrien.favier@ibs.fr) or P. S. (paul.schanda@ibs.fr).

## References

1. Kühlbrandt, W. The resolution revolution. *Science* **343**, 1443–1444 (2014).
2. Cheng, Y. Single-particle cryo-EM - how did it get here and where will it go. *Science* **361**, 876–880 (2018).
3. Tugarinov, V., Choy, W.-Y., Orekhov, V. Y. & Kay, L. E. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 622–627 (2005).
4. Schwieters, C. D. *et al.* Solution structure of the 128 kDa enzyme I dimer from *Escherichia coli* and its 146 kDa complex with HPr using residual dipolar couplings and small-and wide-angle X-ray scattering. *J. Am. Chem. Soc.* **132**, 13026–13045 (2010).
5. Bertelsen, E. B., Chang, L., Gestwicki, J. E. & Zuiderweg, E. R. Solution conformation of wild-type *E. coli* Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8471–8476 (2009).
6. Sprangers, R. & Kay, L. E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**, 618–622 (2007).
7. Religa, T. L., Sprangers, R. & Kay, L. E. Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* **328**, 98–102 (2010).
8. Mas, G. *et al.* Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle. *Sci. Adv.* **4**, eaau4196 (2018).

9. Andrew, E., Bradbury, A. & Eades, R. Nuclear magnetic resonance spectra from a crystal rotated at high speed. *Nature* **182**, 1659 (1958).
10. Warschawski, D. & Loquet, A. Database of protein structures determined by solid-state NMR, [http://www.loquetlab.org/ssnmr\\_struct](http://www.loquetlab.org/ssnmr_struct) (2018).
11. Wang, S. *et al.* Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein. *Nat. Methods* **10**, 1007–12 (2013).
12. Demers, J.-P. *et al.* High-resolution structure of the Shigella type-III secretion needle by solid-state NMR and cryo-electron microscopy. *Nat. Commun.* **5**, 4976 (2014).
13. Sborgi, L. *et al.* Structure and assembly of the mouse ASC inflammasome by combined NMR spectroscopy and cryo-electron microscopy. *Proc. Natl. Acad. Sci.* 201507579 (2015).
14. Gremer, L. *et al.* Fibril structure of amyloid- $\beta$  (1–42) by cryo-electron microscopy. *Science* **358**, 116–119 (2017).
15. Durá, M. A. *et al.* Characterization of a TET-like aminopeptidase complex from the hyperthermophilic archaeon *Pyrococcus horikoshii*. *Biochemistry* **44**, 3477–86 (2005).
16. Borissenko, L. & Groll, M. Crystal structure of TET protease reveals complementary protein degradation pathways in prokaryotes. *J. Mol. Biol.* **346**, 1207–1219 (2005).
17. Amero, C. *et al.* A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. *J. Biomol. NMR* **50**, 229–236 (2011).



18. Mas, G., Crublet, E., Hamelin, O., Gans, P. & Boisbouvier, J. Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins. *J. Biomol. NMR* **57**, 251–262 (2013).
19. Schmidt, E. & Guntert, P. A new algorithm for reliable and general NMR resonance assignment. *J. Am. Chem. Soc.* **134**, 12817–12829 (2012).
20. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).
21. Güntert, P. & Buchner, L. Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR* **62**, 81–95 (2015).
22. Terwilliger, T. C., Adams, P. D., Afonine, P. V. & Sobolev, O. V. A fully automatic method yielding initial models from high-resolution electron cryo-microscopy maps. *bioRxiv* 267138 (2018).
23. Adams, P. D. *et al.* Phenix: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 213–221 (2010).
24. Baker, M. L., Ju, T. & Chiu, W. Identification of secondary structure elements in intermediate-resolution density maps. *Structure* **15**, 7–19 (2007).
25. López-Blanco, J. R. & Chacón, P. IMODFIT: Efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J. Struct. Biol.* **184**, 261–270 (2013).

26. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003).
27. Schwieters, C. D., Kuszewski, J. J. & Marius Clore, G. Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson.* **48**, 47–62 (2006).
28. Afonine, P., Headd, J., Terwilliger, T. & Adams, P. New tool: phenix. real\_space\_refine. *Computational Crystallography Newsletter* **4**, 43–44 (2013).
29. Tang, Y. *et al.* Protein structure determination by combining sparse NMR data with evolutionary couplings. *Nat. Methods* **12**, 751–754 (2015).
30. Morcos, F. *et al.* Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1293–E1301 (2011).
31. Marks, D. S. *et al.* Protein 3d structure computed from evolutionary sequence variation. *PloS one* **6**, e28766 (2011).
32. O'Reilly, F. J. & Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* **1** (2018).
33. Truong, K. & Ikura, M. The use of FRET imaging microscopy to detect protein–protein interactions and protein conformational changes in vivo. *Curr. Opin. Struct. Biol.* **11**, 573–578 (2001).
34. Jeschke, G. Distance measurements in the nanometer range by pulse EPR. *ChemPhysChem* **3**, 927–932 (2002).

35. Reginsson, G. W. & Schiemann, O. Pulsed electron–electron double resonance: beyond nanometre distance measurements on biomacromolecules. *Biochem. J.* **434**, 353–363 (2011).
36. Milikisiyants, S. *et al.* Oligomeric structure of anabaena sensory rhodopsin in a lipid bilayer environment by combining solid-state nmr and long-range deer constraints. *J. Mol. Biol.* **429**, 1903–1920 (2017).
37. Pell, A. J., Pintacuda, G. & Grey, C. P. Paramagnetic nmr in solution and the solid state. *Prog. Nucl. Magn. Reson. Spectrosc.* (2018).
38. Kerfah, R., Plevin, M. J., Sounier, R., Gans, P. & Boisbouvier, J. Methyl-specific isotopic labeling: A molecular tool box for solution NMR studies of large proteins. *Curr. Opin. Struct. Biol.* **32**, 113–122 (2015).
39. Lacabanne, D., Meier, B. H. & Böckmann, A. Selective labeling and unlabeled strategies in protein solid-state NMR spectroscopy. *J. Biomol. NMR* **71**, 141–150 (2018).
40. Gans, P. *et al.* Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew. Chem. Int. Ed.* **49**, 1896 (2010).
41. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–96 (2005).
42. Ludtke, S. J., Baldwin, P. R. & Chiu, W. Eman: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).

43. Scheres, S. H. Relion: implementation of a bayesian approach to cryo-em structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
44. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
45. Estrozi, L. F. & Navaza, J. Ab initio high-resolution single-particle 3d reconstructions: the symmetry adapted functions way. *J. Struct. Biol.* **172**, 253–260 (2010).
46. Jiang, W., Baker, M. L., Ludtke, S. J. & Chiu, W. Bridging the Information Gap : Computational Tools for Intermediate Resolution Structure Interpretation. *J. Mol. Biol.* 1033–1044 (2001).
47. Matthew L. Baker, C. F. H. T. J., Mariah R. Baker & Chiu, W. Gorgon and pathwalking: Macromolecular modeling tools for subnanometer resolution density maps. *Biopolymers* **97**, 655–668 (2012).
48. Pettersen, E. F. *et al.* UCSF chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
49. Schwieters, C. D., Bermejo, G. A. & Clore, G. M. Xplor-NIH for molecular structure determination from NMR and other data sources. *Protein Sci.* **27**, 26–40 (2018).
50. Tian, Y., Schwieters, C. D., Opella, S. J. & Marassi, F. M. High quality NMR structures: a new force field with implicit water and membrane solvation for Xplor-NIH. *J. Biomol. NMR* **67**, 35–49 (2017).

51. Lowther, W. T. & Matthews, B. W. Metalloaminopeptidases: Common functional themes in disparate structural surroundings. *Chem. Rev.* **102**, 4581–4607 (2002).
52. Jiang, J.-S. & Brünger, A. T. Protein hydration observed by x-ray diffraction: solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol.* **243**, 100–115 (1994).

Table 1: Number of distance and dihedral-angle restraints from NMR spectroscopy (top), and structure statistics of the final structure ensemble (20 structures) obtained from all NMR data and the 4.1 Å cryo-EM density map.

Structural restraints (all intra-subunit)	
Total distance restraints (calculation set)	516
short-range ( $ i - j  < 4$ )	203
long-range ( $ i - j  \geq 4$ )	313
unambiguous H-H	441
unambiguous C-C	30
ambiguous H-H	45
Dihedral angle restraints ( $\phi, \psi$ )	544
Ramachandran plot statistics (%)	
most favored regions	$93 \pm 1$
allowed regions	$4 \pm 1$
outliers	$3 \pm 1$
Ensemble backbone precision (Å)	
	0.29

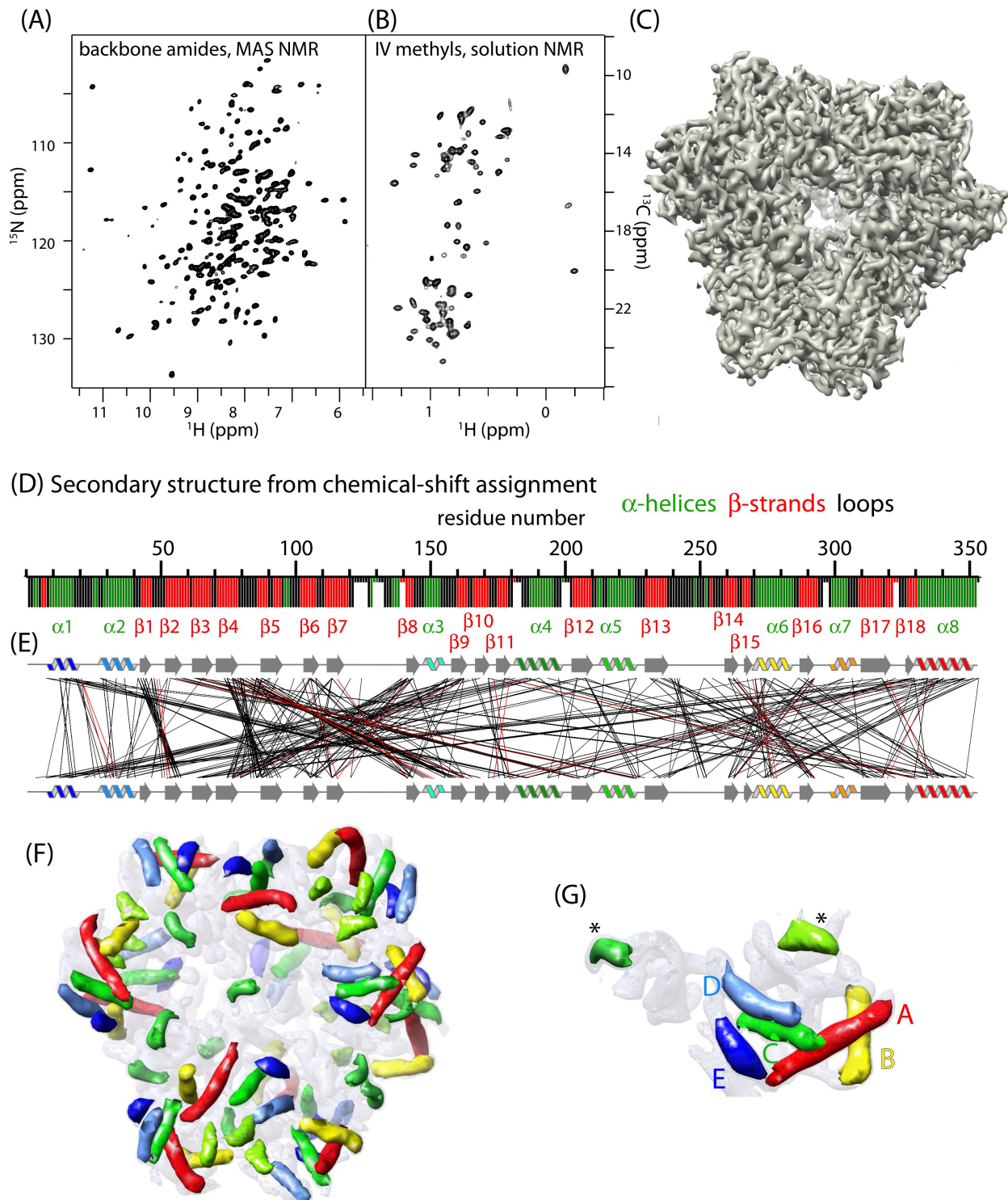


Figure 1: Experimental NMR and EM data of the 468 kDa TET2 assembly. (A) MAS NMR and (B) solution-NMR spectra of TET2, showing high resolution despite the large subunit size. (C) Experimental 4.1 Å resolution cryo-EM electron-density map. (D) Secondary structure of TET2, derived from MAS NMR resonance assignments and the TALOS-N software<sup>20</sup>, shown as a function of the residue number. Residues shown with shorter bars were not assigned, and the secondary structure assignment results from a database approach in TALOS-N. (E) Experimentally detected intra-subunit long-range distance restraints from solution-NMR and MAS NMR, displayed through lines connecting residues in close spatial proximity. Note that part of these distance restraints were spectrally ambiguous, i.e., could be assigned to several atom-atom pairs, and was rendered unambiguous throughout the structure calculation approach (displayed in red). See Table 1 for restraints statistics. All NMR experiments performed in this study and acquisition parameters are listed in Tables S2 to S5. (F)  $\alpha$ -helices detected by the *helixhunter2* software<sup>24,46</sup> in the EM map truncated to 8 Å resolution. Symmetry-related  $\alpha$ -helices are shown in equal colors. Additional  $\beta$ -sheet parts, automatically detected by *gorgon*<sup>47</sup> are shown in Figure S9A. (G) Zoom on one subunit, identified by a clustering analysis (Figure S9). The five longest  $\alpha$ -helices, used for the initial structure calculation steps are labeled with A to E in order of decreasing length (see Table S1).



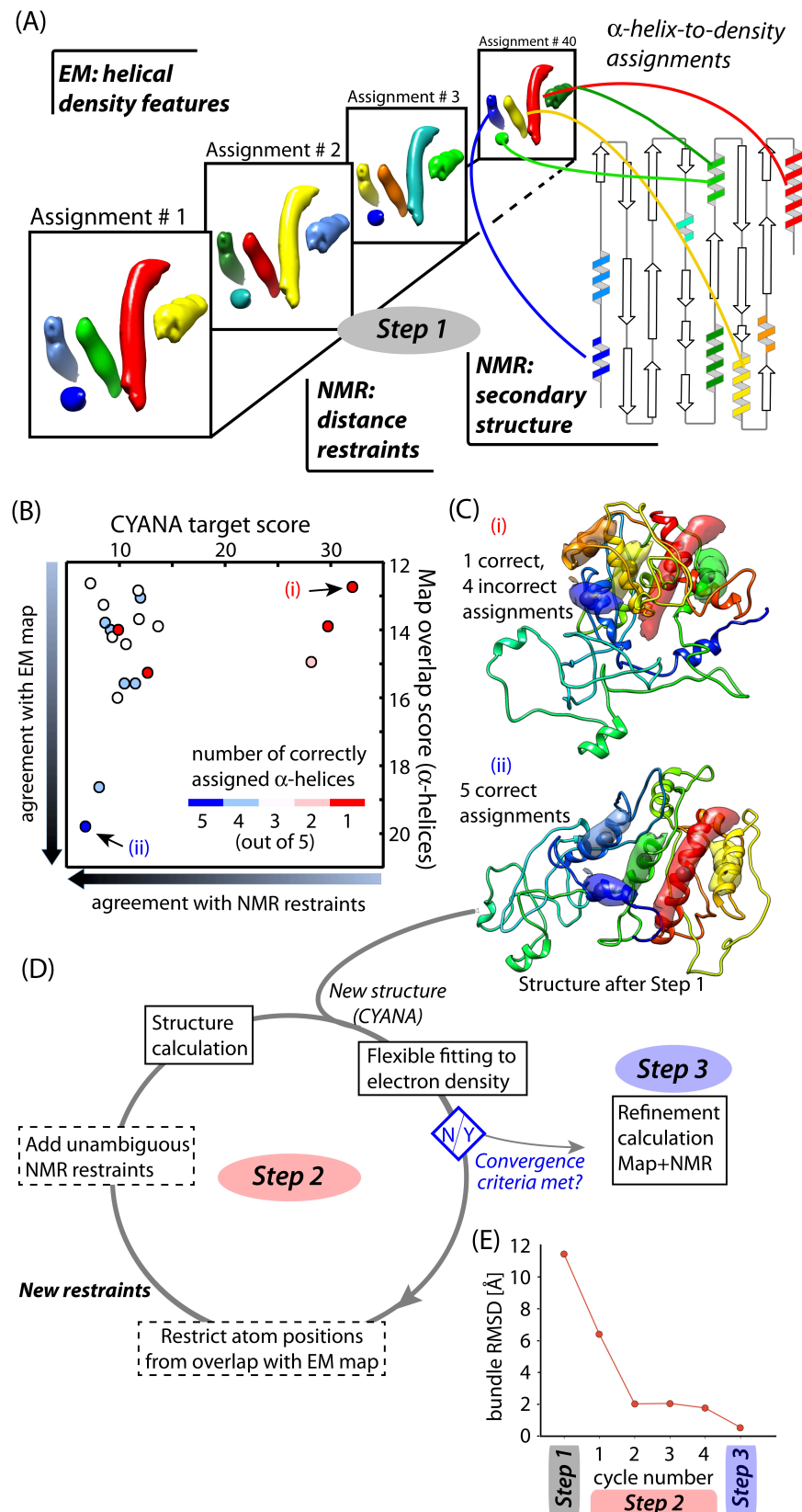


Figure 2: Integrated NMR/EM structure determination approach. (A) In step 1, forty different assignment possibilities of the five longest helical stretches in the sequence (TALOS) to cylindrical (helical) densities ("α-helix-to-density assignments") are used in regular NMR-type structure calculations (CYANA).

(B) Ranking of the twenty solutions from these structure calculations by the CYANA target score and the overlap of the α-helices with the cylindrical density features. Each point represents the average of the 10 lowest-energy structures.

(C) Two cases are shown with incorrect (top) and correct (bottom) assignments, clearly showing that incorrect α-helix-to-density assignment is not compatible with good map overlap. For simplicity, only the lowest-energy structure is shown. The structure with the correct helix-to-density assignment (labeled with (ii)) has a backbone root-mean-square-deviation, RMSD, to the crystal structure of 7.4 Å, and a backbone bundle-RMSD computed from the 20 lowest-energy models of 11.4 Å.

(D) In step 2, the structure of the TET2 monomeric subunit with the correct α-helix-to-density assignment is iteratively refined by flexible fitting into the EM map (truncated at 8 Å resolution), and CYANA calculations with an increasing number of unambiguous NMR restraints and restraints from the fit in the electron density. After convergence, defined by at least three cycles with an RMSD-difference below 10 % (shown in (E)), the structure is refined against the EM map of the entire dodecameric ensemble and all NMR restraints, using XPLOR-NIH, using maps with increasing resolution (8 Å, 6 Å, 4.1 Å; Step 3).

(E) Root-mean-square-deviation (RMSD) of the structures at different steps of the protocol, relative to the mean structure.

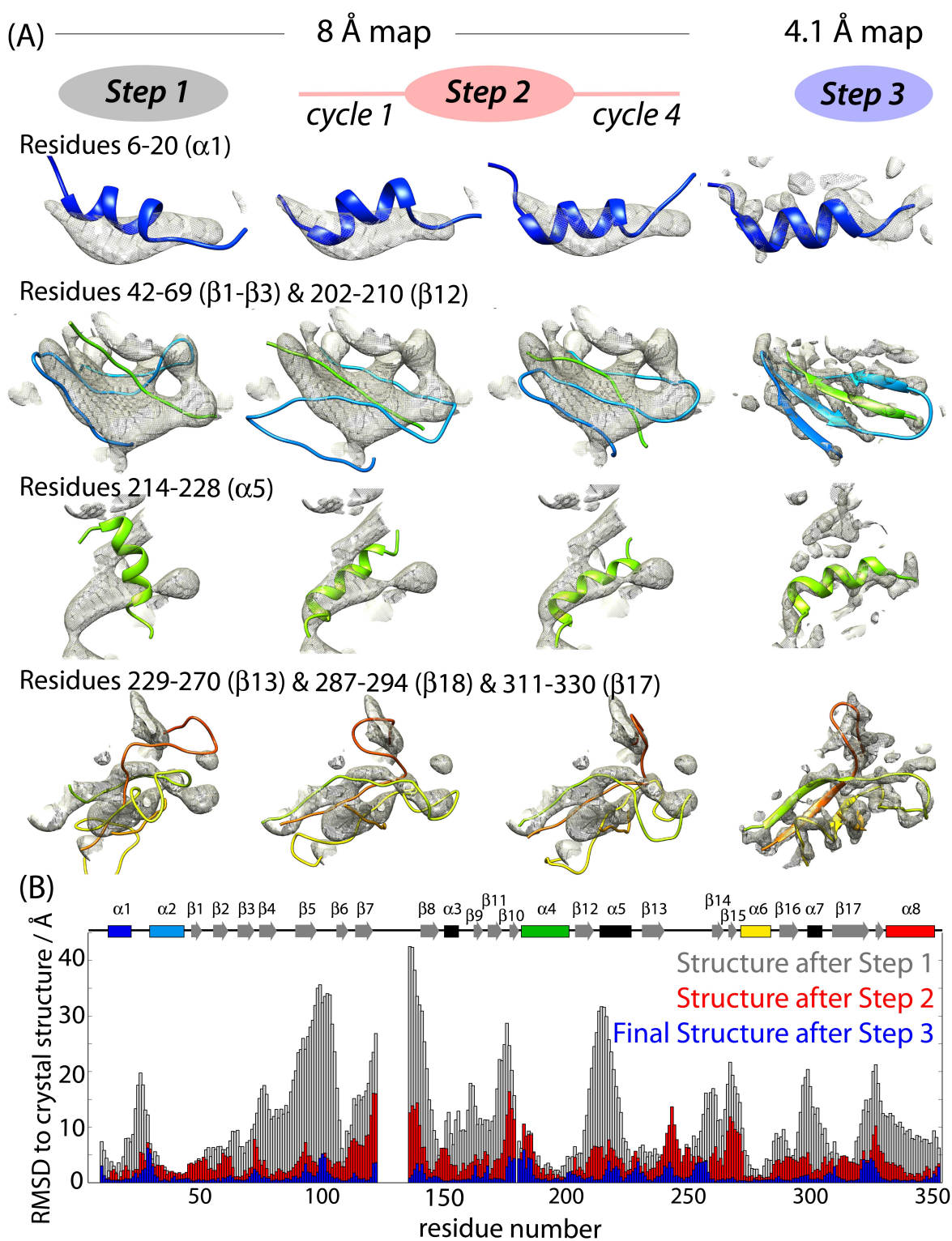


Figure 3: Structure refinement from NMR and EM data. (A) Zoom onto selected parts of the TET2 structure along the different steps of the structure determination process. The 3D EM map used at 8 Å resolution (steps 1 and 2) and 4.1 Å resolution (step 3) is shown as a gray surface. The structures shown correspond to the lowest-energy models generated by CYANA (Steps 1 and 2) or XPLOR-NIH (Step 3). A comprehensive view of the evolution of all structure elements is provided in Figure [S11](#). (B) Residue-wise heavy-atom backbone RMSD relative to the crystal structure throughout different steps of the structure determination protocol.

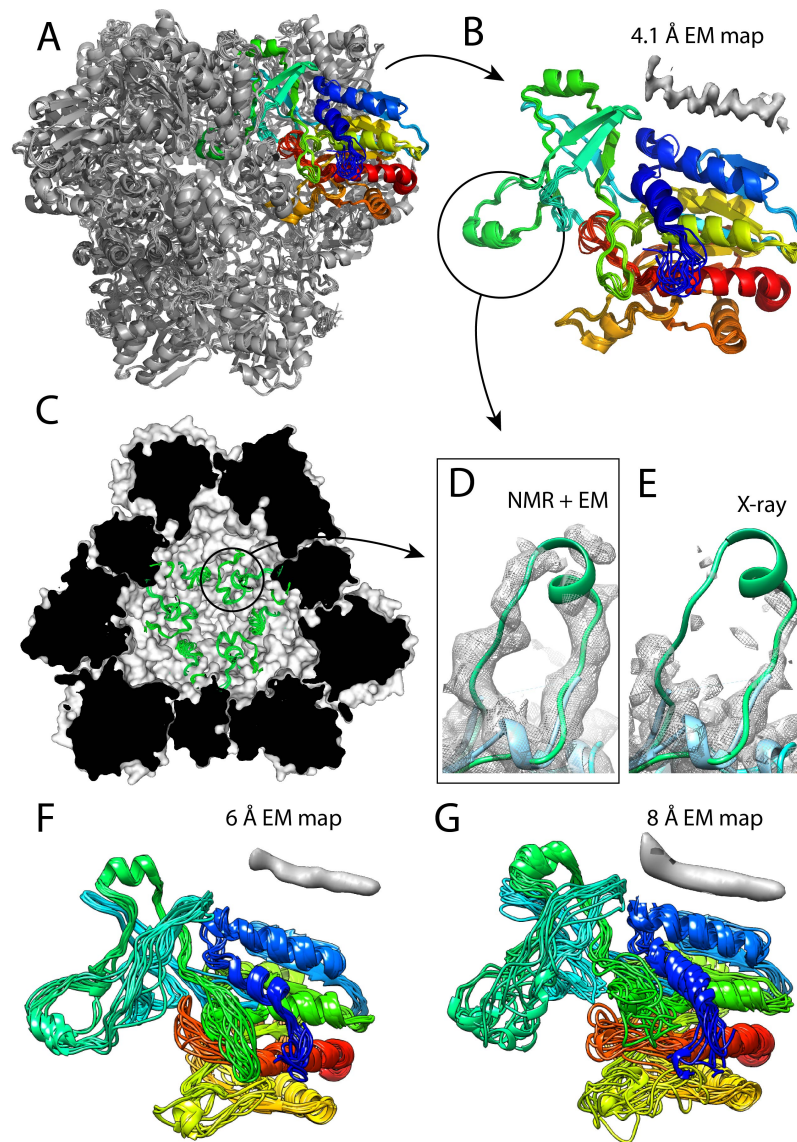


Figure 4: Structure of the TET2 dodecamer obtained from joint refinement of twelve subunits against NMR and EM data. (A) Overall view of the dodecamer, highlighting in color one out of the twelve modeled subunits. (B) Bundle of 10 structures of one monomer, with an accuracy of 0.7 Å (backbone RMSD relative to crystal structure). The loop region, residues 120-138, which has not been modeled in the crystal structure (PDB 1y0r) is encircled in (B), and shown in the view of the interior of the cavity (C) and as a zoom (D). The EM map around the loop region, seen by cryo-EM (D) is of significantly higher intensity than in the 100 K crystal structure (1y0r). The backbone structure in both panels is from the present NMR/EM study. (F, G) Results of NMR/EM structure calculation using only EM data truncated to 6 Å or 8 Å resolution, resulting in backbone RMSD to the crystal structure of 1.8 Å and 2.6 Å, respectively. The inserts above the structures in (B), (F) and (G) represent typical  $\alpha$ -helical densities at the respective resolution levels.

## Supporting Information

### **Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex**

**Diego F. Gauto<sup>1</sup>, Leandro F. Estrozi<sup>1</sup>, Charles D. Schwieters<sup>2</sup>, Gregory Effantin<sup>1</sup>, Pavel Macek<sup>1,6</sup>, Remy Sounier<sup>1,7</sup>, Astrid C. Sivertsen<sup>1</sup>, Elena Schmidt<sup>3</sup>, Rime Kerfah<sup>1,6</sup>, Guillaume Mas<sup>1</sup>, Jacques-Philippe Colletier<sup>1</sup>, Peter Güntert<sup>3,4,5</sup>, Adrien Favier<sup>1</sup>, Guy Schoehn<sup>1</sup>, Jerome Boisbouvier<sup>1</sup>, & Paul Schanda<sup>1</sup>**

<sup>1</sup>Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), 71, avenue des martyrs, F-38044 Grenoble, France.

<sup>2</sup>Center for Information Technology, National Institutes of Health, Bethesda, USA. <sup>3</sup>Institute of Biophysical Chemistry, Goethe University Frankfurt am Main, Frankfurt am Main, Germany.

<sup>4</sup>Laboratory of Physical Chemistry, ETH Zürich, Zürich, Switzerland.

<sup>5</sup>Graduate School of Science, Tokyo Metropolitan University, Tokyo, Japan.

<sup>6</sup>Present address: NMR-Bio, 5 place Robert Schumann, F-38025 Grenoble, France.

<sup>7</sup>Present address: Institut de Génomique Fonctionnelle, CNRS UMR-5203 INSERM U1191, University of Montpellier, F-34000 Montpellier, France



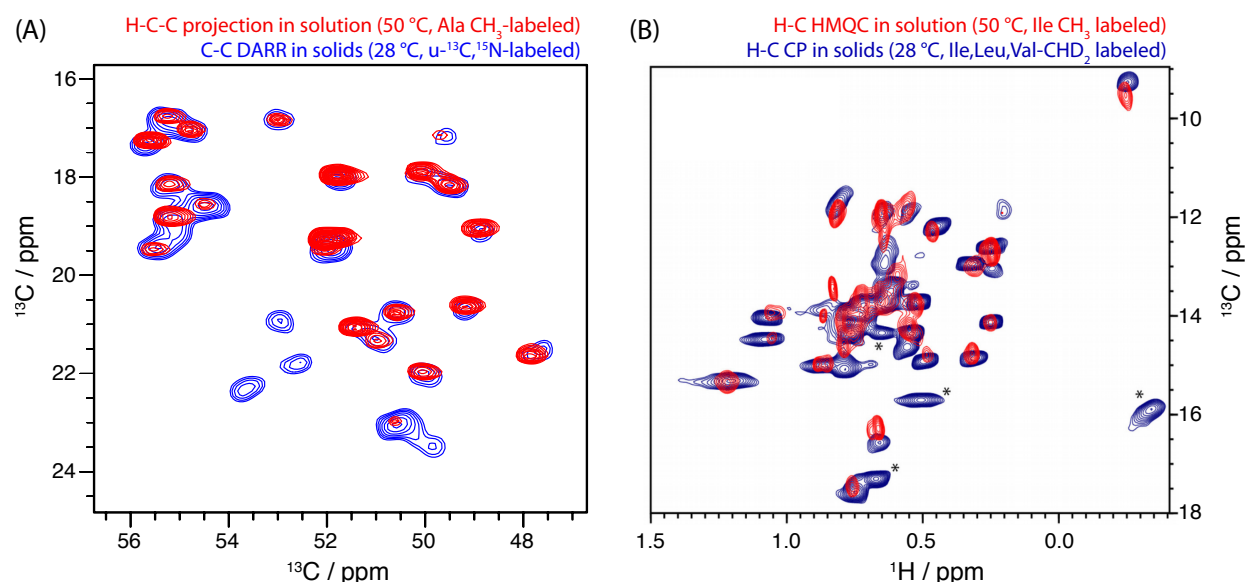


Figure S1: Comparison of solution and MAS NMR spectra of TET2.

(A) Alanine  $\text{C}\alpha$ - $\text{C}\beta$  correlation spectrum. The solution-state NMR spectrum was collected with a deuterated sample labeled with  $^{13}\text{C}^1\text{H}_3$  group and a  $^{13}\text{C}$  spin at the  $\text{C}\alpha$  position. The spectrum is a  $^1\text{H}$  detected  $^1\text{H}$ - $^{13}\text{C}$ - $^{13}\text{C}$  correlation, projected along the  $^1\text{H}$  dimension. The MAS NMR spectrum is a DARR experiment, recorded on uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled sedimented sample, collected at 15 kHz MAS (14.1 T). Note that because of spin relaxation during the transfer steps in the solution-NMR 3D correlation spectrum, several peaks are undetected. The observable peaks match very well the ones in the MAS NMR spectrum.

(B) Comparison of  $^1\text{H}$ - $^{13}\text{C}$  correlation spectra of Ile methyl groups in solution and in an MPD-precipitated TET2 sample. The peaks denoted with an asterisk are resonances from valine sites, which are methyl-labeled in the MAS NMR sample, but not in the solution-NMR sample.



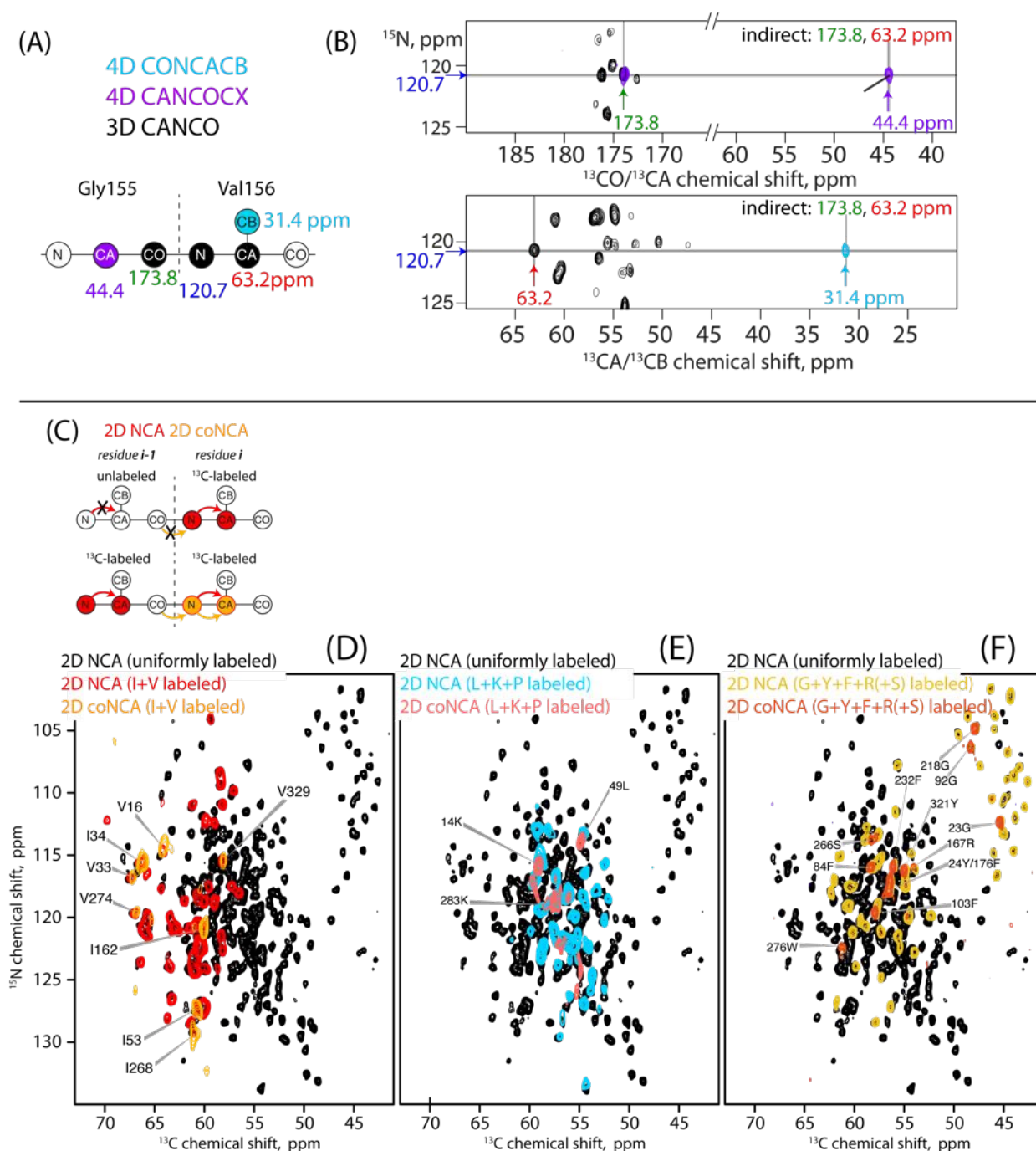


Figure S2: Approach for MAS NMR assignment of TET2 using high-dimensional experiments on uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled samples (A,B) and with the aid of specifically labeled samples (C-F).

(A) Schematic coherence transfer and (B) exemplary spectra showing the assignment strategy based on 4D backbone correlation experiments. The  $\text{CO}_{i-1}$ ,  $\text{N}_i$ ,  $\text{CA}_i$  frequencies observed in the CANCO experiment (black) are well resolved and serve as an unambiguous "root peak" for identifying each site. A fourth frequency dimension is added either towards residue  $i-1$  (CANCOCX) or towards residue  $i$  (CONCACB), thus allowing to unambiguously correlate five frequencies for each residue pair.

(C) Principle of the specific labeling strategy combined with  $^{15}\text{N}$ - $^{13}\text{C}\alpha$  correlation spectra. In the three different samples only certain types of amino acids are labeled with  $^{13}\text{C}$ , and the protein is uniformly  $^{15}\text{N}$ -labeled. In NCA correlation spectra, only residues with  $^{13}\text{C}$  labeling are observable. In co-N-CA correlation spectra, where the initial magnetization is emanating from the carbonyl of the preceding residue, signals are observed only if two consecutive residues are labeled. In (D)-(F) NCA and coNCA correlation spectra from the three differently labeled samples are presented in colors, and compared to a NCA spectrum of a uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled sample. Note that in the G+Y+F+R sample, residual labeling of serine is also obtained through scrambling from glycine. The spectrum of uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled TET was collected at 1 GHz  $^1\text{H}$  Larmor frequency) while all other spectra were recorded at 600 MHz  $^1\text{H}$  Larmor frequency. Residues highlighted with their sequential number are those observed in the coNCA spectra, i.e., for which the preceding residue is also  $^{13}\text{C}$ -labeled. A fully annotated NCA spectrum and amide  $^1\text{H}$ ,  $^{15}\text{N}$  and methyl  $^1\text{H}$ ,  $^{13}\text{C}$  spectra are displayed in Figure S3.

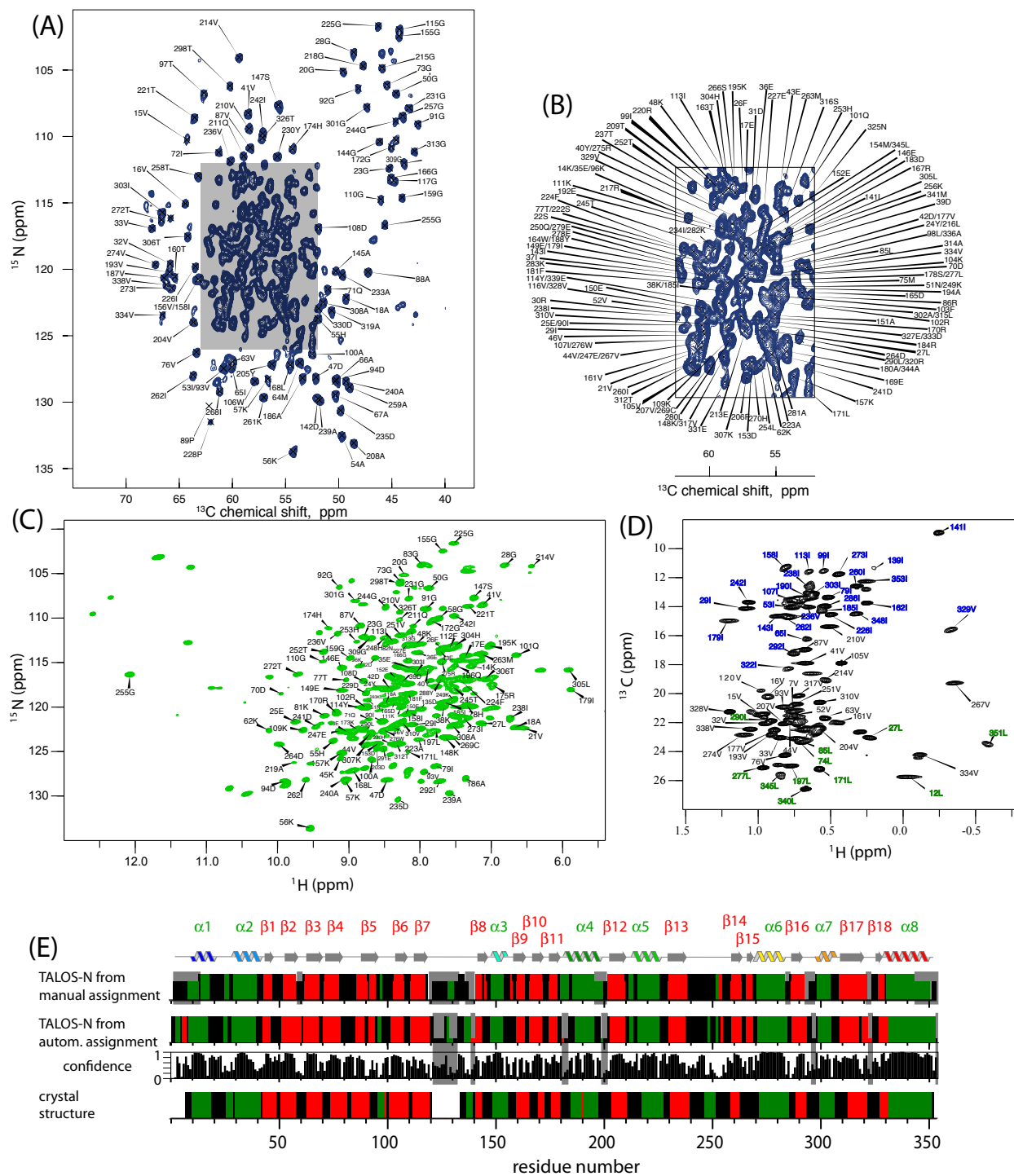


Figure S3: Assigned 2D MAS NMR correlation spectra of TET2. (A) NCA correlation spectrum ( $^{13}\text{C}$ -detected) of uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled TET2 recorded at 1000 MHz  $^1\text{H}$  Larmor frequency, and a zoom of the central part (B).

(C)  $^1\text{H}$ -detected amide  $^1\text{H}$ - $^{15}\text{N}$  correlation spectrum of uniformly  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled TET2.

(D)  $^1\text{H}$ -detected methyl  $^1\text{H}$ - $^{13}\text{C}$  MAS NMR correlation spectrum of u- $[\text{}^2\text{H}, \text{}^{15}\text{N}]$ -Ile $^{\delta 1}$ , Leu $^{\text{proS}}$ , Val $^{\text{proS}}$ - $[\text{}^{13}\text{CHD}_2]$ -labeled TET2 at 600 MHz  $^1\text{H}$  Larmor frequency (14.1 T). (The pro-S position is also referred to as  $\gamma 2$  (in Val) and  $\delta 2$  (in Leu).) Ile, Val and Leu methyl groups are annotated in blue, black and green, respectively.

(E) Secondary structure in TET2 derived from MAS NMR chemical shift assignments using the program TALOS-N<sup>20</sup>, and comparison to secondary structures in the crystal structure. Loop/coil regions are shown in black,  $\alpha$ -helices in green and  $\beta$ -strands in red. Grey bars in the background denote residues for which no backbone assignment was obtained, comprising the long flexible loop from residues 120-133. Note that for residues where no/insufficient chemical-shift assignments are available, TALOS-N uses a database approach to propose secondary structures; these residues are shown as shorter bars in the TALOS-N plot. For comparison, the results from the manual and automatic (FLYA) assignments are shown. The confidence level of the TALOS-N secondary structure predicting is shown for the FLYA data set. The lowest shows the secondary structures as obtained from the crystal structure (PDB entry 1Y0R) and the DSS algorithm, determined with pymol. Missing bars denote residues not modeled in the crystal structure.

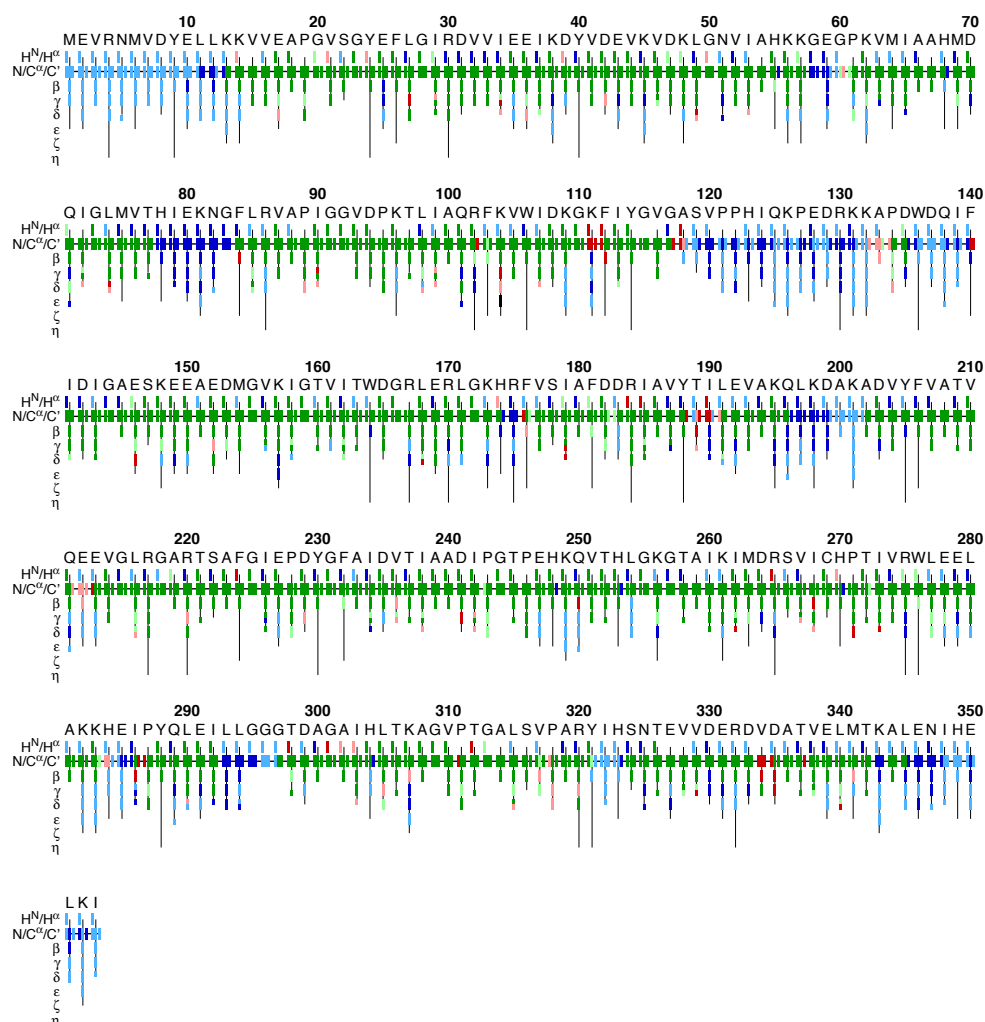


Figure S4: Automatic assignment of MAS NMR spectra of TET2 using the program FLYA.<sup>[19]</sup> The program used the following spectra for assignment: NCA, NCO, NCOCX, CANCO, NCACX, NCACB, NcoCACB, CONCACB, CANCOCX, NCOCX(GRYF), NCOCX(ILV), NCOCX(LKP), NCACX(GRYF), NCACX(ILV), NCACX(LKP), hCANH, hCONH, hcoCAcoNH, CCCdarrdarr(ILV), CCC, caNCO(GRYF), coNCA(GRYF), coNCA(ILV), coNCA(LKP). The color code reflects the assignment for each atom, and compares it to the manual assignment, as follows. Green: Automatic and manual assignment agree; red: automatic assignment differs from manual assignment; blue: no manual assignment available. Dark colors (green, blue, red) denote atoms for which the FLYA assignment has converged, i.e., the same assignment is found in more than 80% of the 20 independent runs, while light colors denote atoms for which FLYA has not converged, i.e., the results have less confidence. The row labeled H<sup>N</sup>/H<sup>α</sup> shows for each residue HN on the left. The N/C<sup>α</sup>/C' row shows for each residue the N, C<sup>α</sup>, and C' assignments from left to right. The rows labeled with β-η show the side-chain assignments for the heavy atoms. In the case of branched side chains, e.g. Val and Leu, the corresponding row is split into an upper part for one branch and a lower part for the other branch.

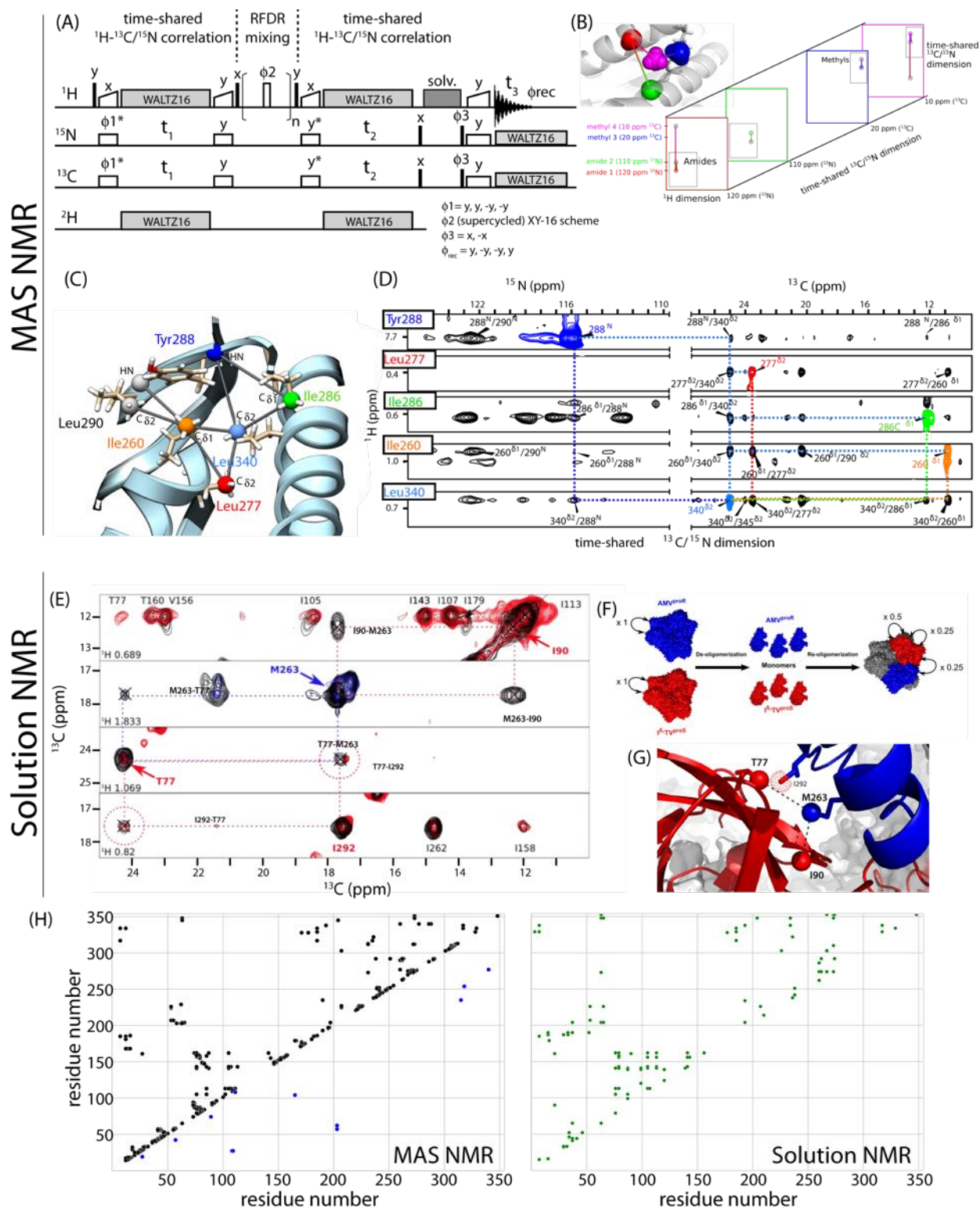




Figure S5: Distance restraints from MAS NMR and solution-NMR experiments.

(A) Pulse sequence used for the simultaneous measurement of  $^1\text{H}$ - $^1\text{H}$  spatial proximities between amide and methyl moieties under fast-MAS conditions, for application to deuterated, amide-protonated and methyl  $^{13}\text{CHD}_2$  labeled samples. Filled and open narrow rectangles denote  $90^\circ$  and  $180^\circ$  pulses, respectively, and wide open rectangles denote cross-polarization steps, applied with a linear ramp on the  $^1\text{H}$  channel. The RFDR mixing, of a total duration of 5 ms, is comprised of a train of  $n$  equally-spaced pulses where the centers of two consecutive pulses are separated by the MAS rotor period; the pulse phase is alternated according to the XY16 scheme. The cross-polarization steps simultaneously transfer polarization from amide- $^1\text{H}$  to  $^{15}\text{N}$  and from the methyl- $^1\text{H}$  to the methyl- $^{13}\text{C}$  (out and back). Pulse phases are denoted above the respective symbols. Quadrature detection in the indirect dimensions ( $t_1$ ,  $t_2$ ) was performed by varying the phase of the  $^{13}\text{C}$  and  $^{15}\text{N}$  cross-polarization irradiation prior to the chemical-shift (denoted by an asterisk) according to the States-TPPI scheme. Decoupling of spin-spin couplings during direct and indirect dimensions and the solvent-suppression scheme ("solv.") are denoted by grey rectangles.

(B) Schematic representation of the outcome of this experiment, showing connections between two amide sites, between two methyl sites and between an amide and a methyl site.

(C) Experimental distance restraints from the RFDR experiment shown on part of the TET2 structure.

(D) Excerpts from the 3D RFDR MAS NMR spectrum, showing the cross-peaks supporting the distance restraints shown in (C).

(E) Subunit-specific labeling of TET2 to filter inter-subunit distance information. Example strips from 3D HMQC-NOESY-HMQC spectra recorded on either u- $[\text{}^2\text{H}, \text{}^{15}\text{N}]$ , Ile $^{\delta 1}$ , Thr $^{\gamma 2}$ , Val $^{\text{proS}}$ - $[\text{}^{13}\text{CH}_3]$  (red), u- $[\text{}^2\text{H}, \text{}^{15}\text{N}]$ , Met $^{\epsilon}$ , Ala $^{\beta}$ , Val $^{\text{proR}}$ - $[\text{}^{13}\text{CH}_3]$  (blue) or a sample in which these two differently labeled subunit types were mixed in a 1:1 ratio (black).

(F) Labeling scheme for these spectra. The numbers reflect the relative cross-peak intensities resulting from the stochastic mixing scheme.

(G) Illustration of inter-subunit distance restraints detected between M263 in one subunit and T77 and I90 in another subunit. An additional methyl group from which an intra-subunit contact is observed in the spectra in (C) is shown with a dotted sphere. Note that the inter-subunit distance restraints have not been used directly in this work, but we have only used this information for excluding inter-subunit restraints from the calculations (see Methods). A total of only eight inter-subunit distances has been identified, and this set of experiments might not be crucial in the presented approach.

(H) Residue contact maps showing the restraints from the  $^1\text{H}$ - $^1\text{H}$  MAS NMR RFDR experiment connecting amide and methyl sites (black), the  $^{13}\text{C}$  DARR experiment on LKP-labeled TET2 (blue), and the solution-state NOESY experiments (green). The eight inter-subunit restraints have been removed from this graph.

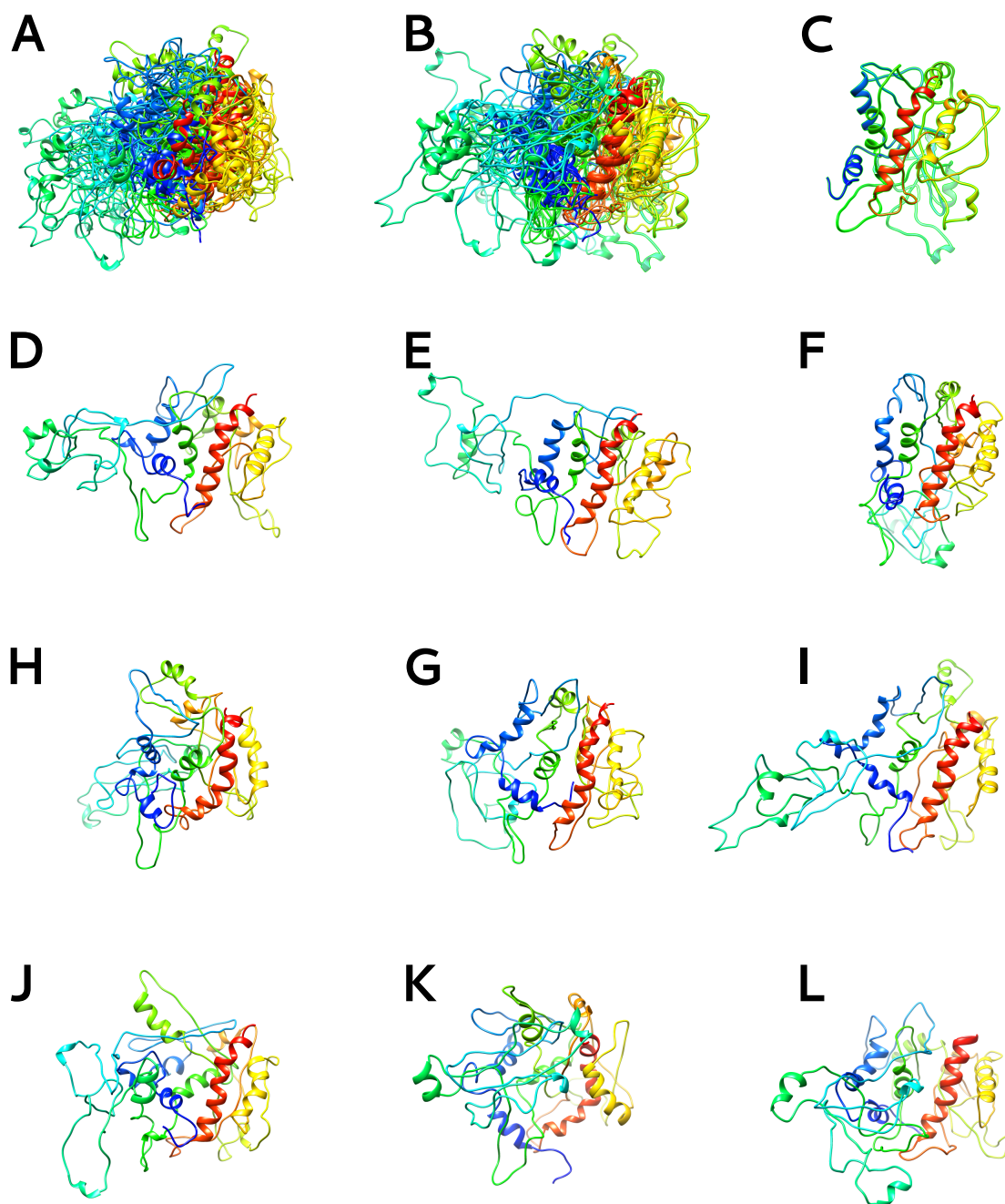


Figure S6: Structure calculation of the monomer of TET2 using only the liquid and solid-state NMR derived restraints. Figure (A) represents the NMR ensemble of 10 structures superimposed on their secondary structure elements (backbone RMSD:  $10.5 \pm 1.6$  Å). Figure (B) shows the same NMR structure ensemble superimposed on the largest helix  $\alpha 8$ , each structure of this ensemble are displayed individually on figures (C) to (L).



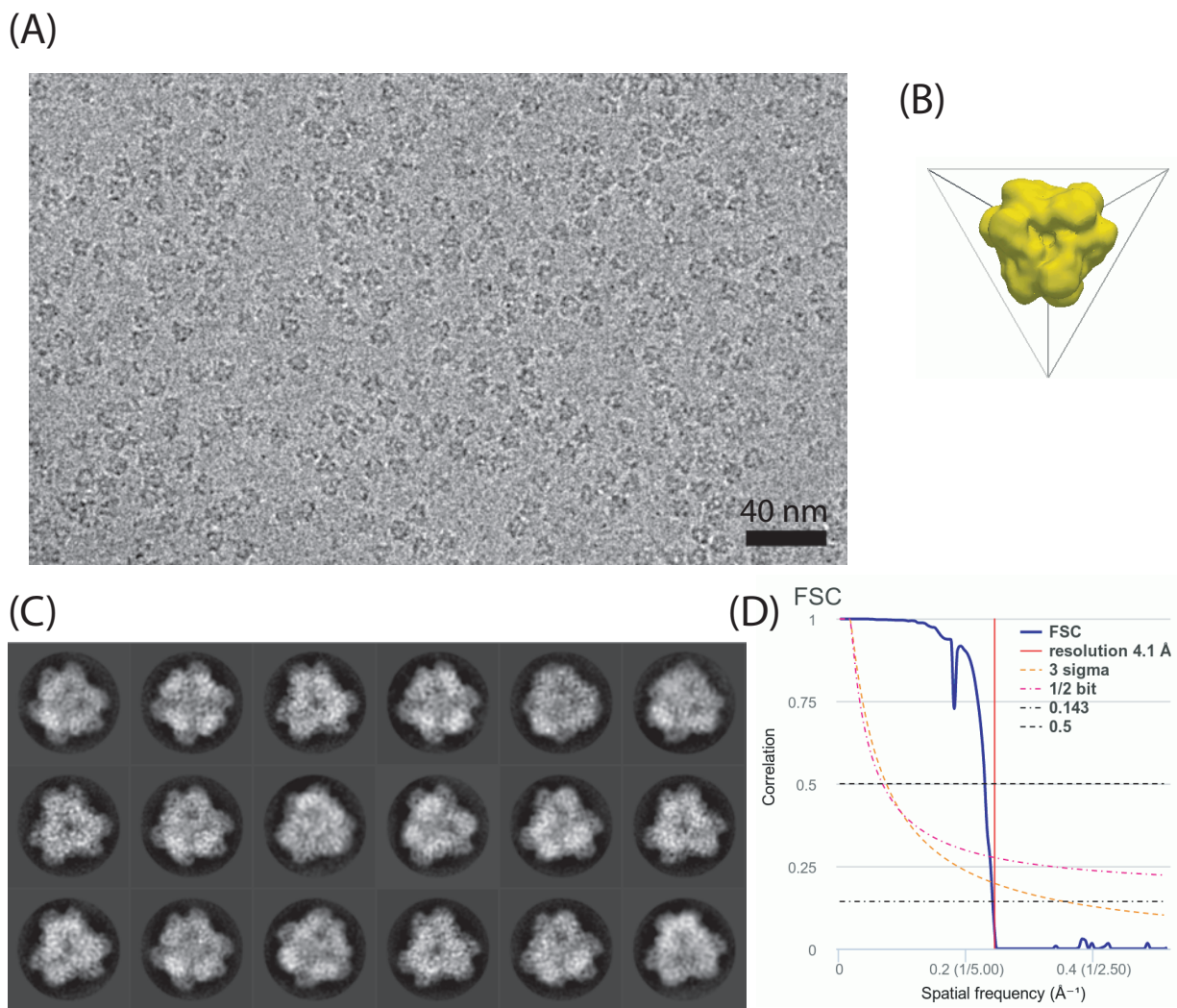


Figure S7: (A) Typical cryo-electron micrograph showing TET particles (darker than the background densities) in random orientations. (B) Used initial low resolution 3D model obtained by imposition of tetrahedral symmetry. (C) Best eighteen 2D class averages obtained from 27,130 particles. (D) Fourier Shell Correlation plot for the presented 3D cryo-EM reconstruction.

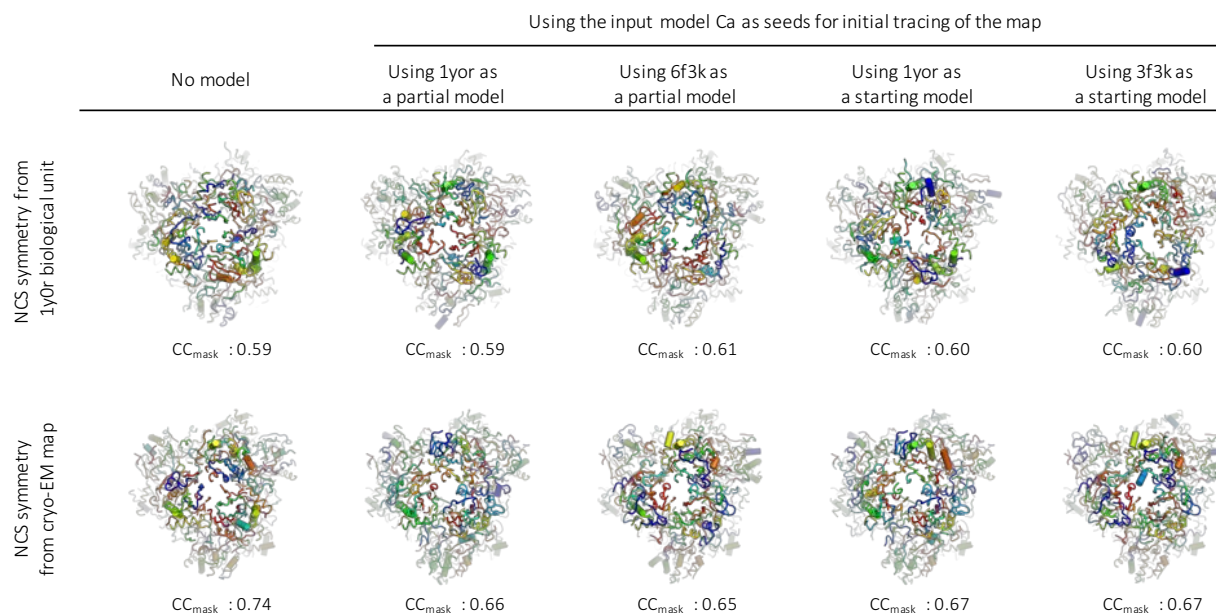


Figure S8: Attempts to determine the structure from the EM map only. Direct rebuilding from the 4.1 Å cryo-EM map fails to produce interpretable structures, regardless of the symmetry that is imposed (or not) for model-building, and of whether (or not) a partial or a starting model is provided. Briefly, we used phenix.map\_to\_model (10 rebuild cycles) to attempt rebuilding the structure directly from the cryo-EM map (4.1 Å resolution) using either no prior information (no model), or the X-ray (1yor) or the NMR+cryo-EM (6f3k, this paper) structures as starting coordinates. We tested inclusion of these models either as "partial" or "starting" models, and in all cases, imposed the use of the input-model C $\alpha$ -atoms as seeds for the initial tracing of the map. Rebuilding was attempted with (upper row) or without (lower row) imposing the known symmetry of the particle. In both case, phenix.map\_to\_model was able to trace the map, but detection of the NCS symmetry from the map (lower panel), as opposed to inferred from the X-ray pdb structure (upper panel), drastically improved the quality of the model. At the map resolution, however, phenix.model\_to\_map was not able to interpret the electron density in terms of secondary structure and/or sequence at least not in 10 rebuild cycles. Arguably, the traced model produced by phenix.map\_to\_model would have been interpretable by an experienced crystallographer or electron microscopist, but reaching the quality of 6f3k would have been very challenging.

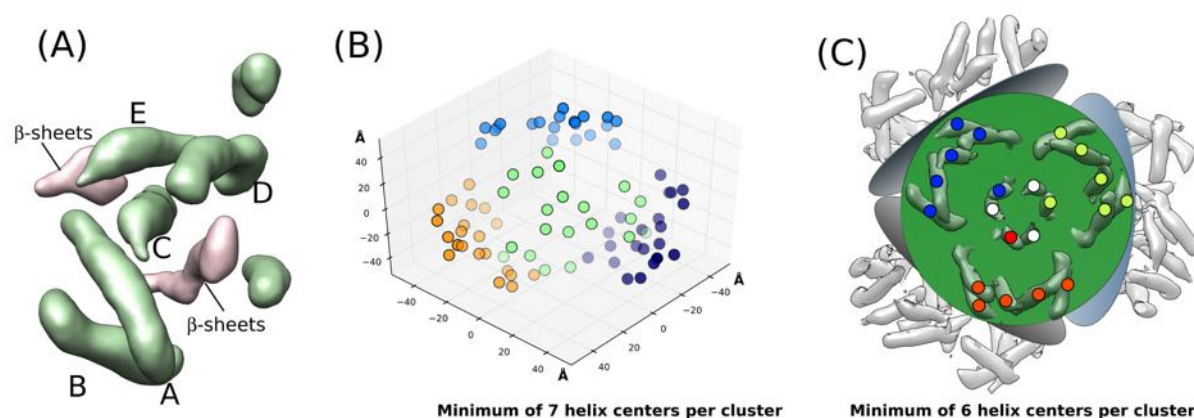


Figure S9: Automatic detection and clustering of secondary structures. (A) Secondary structures detected by *gorgon*<sup>47</sup> from the 8 Å EM map of TET2, zooming on one TET2 subunit.  $\alpha$ -helices and  $\beta$ -sheet regions are depicted in green and pink, respectively. Note that in this study we have only used five  $\alpha$ -helices, labeled with A to E, and we have not used the information about the location of the  $\beta$ -sheets. Including these regions is straightforward and may prove useful in other applications, with proteins containing primarily  $\beta$ -sheets.

(B,C) Cluster analysis of the  $\alpha$ -helices from the EM map, aimed at identifying the  $\alpha$ -helices belonging to the same monomer. This automatic distance-based clustering analysis used the python library *sklearn.cluster*, and needs as input (i) the coordinates of the points in space to be clustered – here, the centers of the 84 automatically detected  $\alpha$ -helices –, (ii) the minimum number of elements in each cluster and (iii) the maximum distance between the points belonging to a given cluster. Note that no information about the symmetry or the number of subunits (number of clusters) is provided.

In (B), the minimum number of elements belonging to the same cluster was set to 7. The points distributed in space, representing the centers of the  $\alpha$ -helices in the dodecamer, are color-coded according to the group/cluster to which they are assigned. While this clustering does not retrieve the monomer subunits, it detects clusters of 21 points, which correspond to trimers of TET2.

In (C), this analysis was repeated, setting the minimum number of elements per cluster to 6. In this case, the six points assigned to each cluster (shown in blue, red or green) indeed correspond to helix centers of a given subunit. The additional white dots represent the helix centers that could not be assigned to any cluster using the distance-based clustering algorithm. The green plane is shown for facilitating visualization of one trimer within the dodecamer. In Step 1 of our structure calculation protocol we have used the five  $\alpha$ -helices within these clusters which are closest in space to each other, omitting the sixth helix.

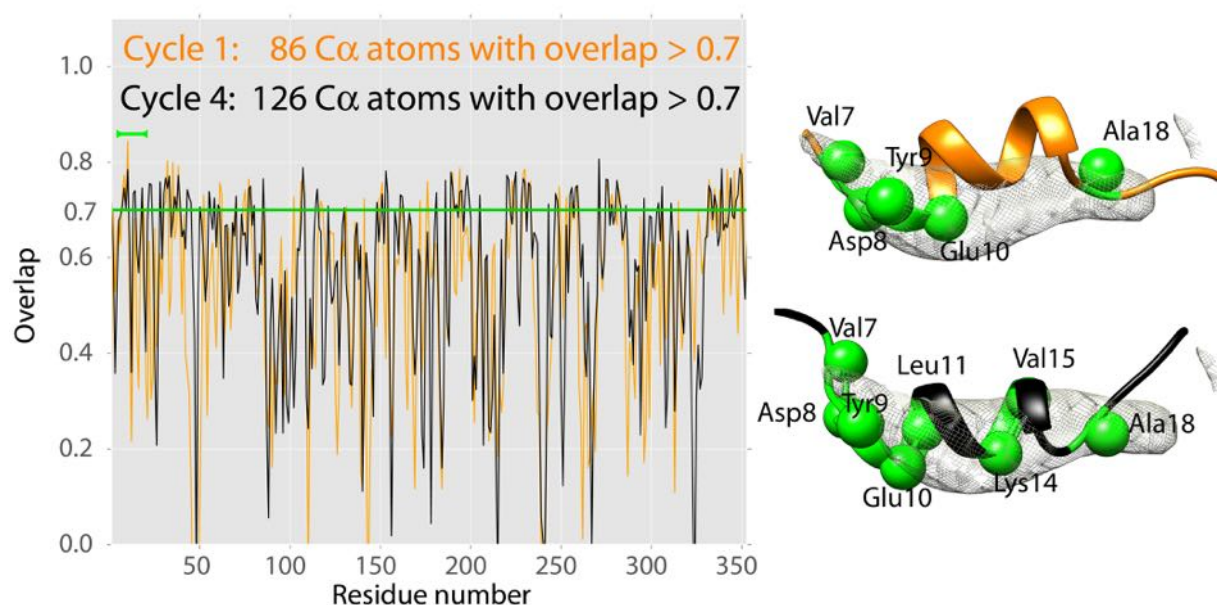


Figure S10: Illustration of the procedure through which, in Step 2 of this approach, residues were restrained to real-space coordinates, based on their overlap with the electron density map. Briefly, after flexible fitting with iModFit the structure was compared to the 8 Å electron density map to evaluate how well different parts of the chain reside inside the density. To this end, the *molmap* and *measure correlation* modules in UCSF Chimera were used to compute the overlap of an in silico map, constructed from the backbone atoms of a given residue, with the experimental electron density. The correlation between these two maps, a value ranging from 1 (atom is fully inside density) to 0 (atom outside density) was used as a criterion whether the residue resides within the map or not. If the value was above 0.7 for a given residue, the C $\alpha$  atom of this residue was then kept in place in the subsequent CYANA calculation by adding restraints from this atom to the other atoms that are fixed in space (using a tolerance of 0.5 Å.)

(A) Residue-wise plot of the overlap score in the first and fourth iteration of Step 2. The green bar on the top left indicates the residues which are shown in (B).

(B) Example of the overlap in the first and fourth iteration. C $\alpha$  atoms with an overlap score above 0.7 are indicated with green spheres and were constrained in the space with a tolerance of  $\pm 2\text{Å}$ .



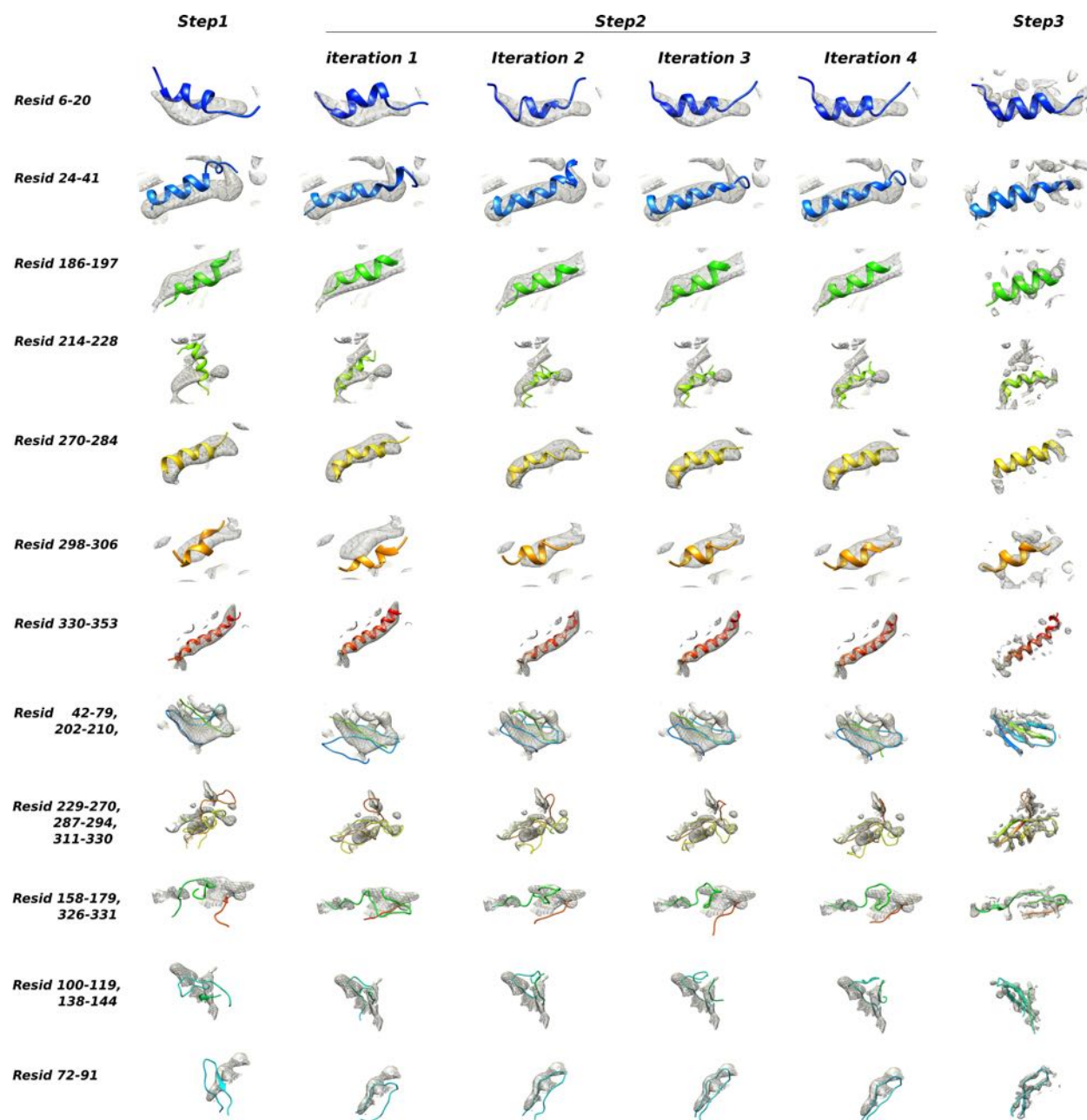


Figure S11: Evolution of the structure of TET2 along the various steps of the integrated NMR/EM structure calculation approach, denoted as Step 1 to 3 in Figure 2. The structures shown correspond to the lowest-energy models generated by CYANA (Steps 1 and 2) or XPLOR-NIH (Step 3).

(A) $\alpha$ helices cryo-EM	Length (cryo-EM)	$\alpha$ helices NMR
A	30.8 Å	$\alpha 8$
B	20.2 Å	$\alpha 6, \alpha 4$
C	21.7 Å	$\alpha 6, \alpha 4$
D	18.1 Å	$\alpha 6, \alpha 4, \alpha 5, \alpha 2, \alpha 1, \alpha 7, \alpha 3$
E	13.1 Å	$\alpha 5, \alpha 2, \alpha 1, \alpha 7, \alpha 3$
(B) Residues placed in centers of the $\alpha$ -helices and estimated lengths in Å		
$\alpha 1$	Lys 13	13.0 Å
$\alpha 2$	Ile 34	13.0 Å
$\alpha 3$	Ala 151	8.6 Å
$\alpha 4$	Leu 191	15.8 Å
$\alpha 5$	Arg 220	13.0 Å
$\alpha 6$	Leu 277	18.7 Å
$\alpha 7$	Ile 303	8.6 Å
$\alpha 8$	Met 341	31.7 Å
(C) Distance restraints corresponding to the correct helix-to-density assignment		
Met 341 C $\alpha$	Leu 277 C $\alpha$	$10.84 \pm 0.50$ Å
Met 341 C $\alpha$	Leu 191 C $\alpha$	$10.46 \pm 0.50$ Å
Met 341 C $\alpha$	Ile 34 C $\alpha$	$20.36 \pm 0.50$ Å
Met 341 C $\alpha$	Lys 13 C $\alpha$	$19.24 \pm 0.50$ Å
Leu 277 C $\alpha$	Leu 191 C $\alpha$	$18.30 \pm 0.50$ Å
Leu 277 C $\alpha$	Ile 34 C $\alpha$	$28.11 \pm 0.50$ Å
Leu 277 C $\alpha$	Lys 13 C $\alpha$	$28.92 \pm 0.50$ Å
Leu 191 C $\alpha$	Ile 34 C $\alpha$	$10.43 \pm 0.50$ Å
Leu 191 C $\alpha$	Lys 13 C $\alpha$	$12.49 \pm 0.50$ Å
Ile 34 C $\alpha$	Lys 13 C $\alpha$	$12.52 \pm 0.50$ Å

Table S1: (*Preceding page*) Details about the helix-to-density assignments. The number of helix-to-density assignments for eight  $\alpha$ -helices detected in the sequence, placed into the seven  $\alpha$ -helical densities detected in the EM map (i.e., leaving out one helix in each combination), ignoring the helix polarity, can be calculated as  $8! = 40320$ . Including the helix polarity increases this number to  $8! \times 2^7 \approx 5 \cdot 10^7$ . Considering the possibility that the seven densities may not be ascribed unambiguously to a given subunit further increases this number. We solved this latter point (identifying densities of one subunit) by a cluster analysis (see Figure S9) and retaining only the five longest helical densities, which are also closest in space. We explicitly disregarded the helix polarity, by using only the center of the helices in the CYANA calculation, letting the rotation free. The number is furthermore greatly reduced by comparing the lengths of  $\alpha$ -helices along the sequence with the helical densities.

(A) This table shows the lengths (middle column) of the helical densities from the EM map (denoted as A-E as shown in Figure 1G, left column), and the  $\alpha$ -helices along the sequence (right column, see Figure 1D) that may match the lengths of these densities. Hereby, we have considered that a helix has a length of 1.4 Å per residue. In selecting the possible helix-to-density assignments, we applied a tolerance of  $\pm 2$  amino acids (corresponding to  $\pm 2.8$  Å). As a result, only  $1 \times 2 \times 5 \times 4 = 40$  helix-to-density assignments remain that are checked by structure calculations.

(B) Identity of the residues assumed to be in the centers of the  $\alpha$ -helices, as obtained from the NMR assignments and TALOS-N. The lengths of the helices in Å was estimated from the length in number of residues, as obtained by TALOS-N, assuming a canonical translation length per residue of ca. 1.45 Å.

(C) Practical implementation of the helix-helix distance restraints, shown here for the correct helix-to-density assignment. These distance restraints were used in CYANA, alongside the other NMR distance restraints and backbone dihedral-angle restraints.

Table S2: Acquisition parameters of MAS NMR spectra for resonance assignment of protonated,  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled TET2 in 3.2 mm rotors. CP transfer steps between  $^{15}\text{N}$  and  $^{13}\text{C}$  or  $^{13}\text{C}$  RF carrier either at the  $\text{C}^\alpha$  (54 ppm) or CO (175 ppm) frequency. The MAS frequency is denoted in kHz. All spectra at 14.1 T were acquired on a Agilent spectrometer; the three spectra (two 2Ds and one 3D) at 23.5 T were recorded on a Bruker Avance spectrometer. The two experiments denoted with "long CP" used a longer  $^1\text{H}$ - $^{15}\text{N}$  cross-polarization to promote proline signals. The column "peaks" is the number of detected peaks used for the automatic assignment in FLYA.

Experiment	Dim. Bo/T	Transfer 1	Transfer 2	Transfer 3	Transfer 4	direct dimension spectr. width/kHz points/ $t_{\text{max}}$	ind. dim. 1 spectr. width/kHz points/ $t_{\text{max}}$	ind. dim. 2 spectr. width/kHz points/ $t_{\text{max}}$	ind. dim. 3 spectr. width/kHz points/ $t_{\text{max}}$	NS d1/s	# peaks
NCA	2D 23.5 18	CP H-N, 0.9 ms $^1\text{H}$ : 60 kHz (ramp) $^{15}\text{N}$ : 41 kHz	CP N-CA, 4 ms $^{15}\text{N}$ : 12 kHz $^{13}\text{C}$ : 6 kHz	-	-	$^{13}\text{C}$ 75 kHz 2020/27 ms	$^{15}\text{N}$ 3.65 kHz 80/21.9 ms	-	-	32 1.5 s 5.5 h	233
NCO	2D 23.5 18	CP H-N, 0.9 ms $^1\text{H}$ : 60 kHz (ramp) $^{15}\text{N}$ : 41 kHz	CP N-CA, 4 ms $^{15}\text{N}$ : 12 kHz $^{13}\text{C}$ : 6 kHz	-	-	$^{13}\text{C}$ 75 kHz 2020/27 ms	$^{15}\text{N}$ 3.65 kHz 80/21.9 ms	-	-	8 1.5 s 1.5 h	264
NcoCX	2D 14.1 15	CP H-N, 8 ms $^1\text{H}$ : 67 kHz (ramp) $^{15}\text{N}$ : 51 kHz	CP NCO, 5.5 ms $^{15}\text{N}$ : 11 kHz $^{13}\text{C}$ : 3.5 kHz (ramp)	C-C DARR, 130 ms $^1\text{H}$ RF: 15 kHz	-	$^{13}\text{C}$ 50 kHz 1500/15 ms	$^{15}\text{N}$ 3 kHz 64/13.7 ms	-	-	496 1.8 s 41 h	118
NcaCX	2D 14.1 15	CP H-N, 8 ms $^1\text{H}$ : 67 kHz (ramp) $^{15}\text{N}$ : 51 kHz	CP NCO, 5.5 ms $^{15}\text{N}$ : 11 kHz $^{13}\text{C}$ : 3.5 kHz (ramp)	C-C DARR, 80 ms $^1\text{H}$ RF: 15 kHz	-	$^{13}\text{C}$ 50 kHz 1500/15 ms	$^{15}\text{N}$ 4.6 kHz 64/13.7 ms	-	-	496 1.8 s 39.7 h	91
NCOCX	3D 14.1 15	CP H-N, 1.1 ms $^1\text{H}$ : 74 kHz (ramp) $^{15}\text{N}$ : 58 kHz	CP NCO, 5 ms $^{15}\text{N}$ : 12 kHz $^{13}\text{C}$ : 3.5 kHz (ramp)	C-C DARR, 50 ms $^1\text{H}$ RF: 15 kHz	-	$^{13}\text{C}$ 40.3 kHz 1612/20 ms	$^{15}\text{N}$ 3 kHz 32/10.3 ms	$^{13}\text{C}$ 3.5 kHz 32/8.8 ms	-	32 1.88 s 71 h	1172
NCACX	3D 23.5 18	CP H-N, 0.9 ms $^1\text{H}$ : 58 kHz (ramp) $^{15}\text{N}$ : 50 kHz	CP N-CA, 4 ms $^{15}\text{N}$ : 14 kHz $^{13}\text{C}$ : 4 kHz (ramp)	C-C DARR, 80 ms $^1\text{H}$ RF: 17 kHz	-	$^{13}\text{C}$ 75 kHz 2264/30 ms	$^{15}\text{N}$ 3.65 kHz 70/18.9 ms	$^{13}\text{C}$ 9 kHz 68/7.4 ms	-	8 1.95 s 81 h	1574
NCACB	3D 14.1 15	CP H-N, 1.1 ms $^1\text{H}$ : 63 kHz (ramp) $^{15}\text{N}$ : 46 kHz	CP N-CA, 5 ms $^{15}\text{N}$ : 12 kHz $^{13}\text{C}$ : 3.5 kHz (ramp)	CA-CB DREAM, 4 ms $^{13}\text{C}$ avg RF: 7 kHz tan ramp (62 to 138%)	-	$^{13}\text{CO}$ 40.3 kHz 1612/20 ms	$^{15}\text{N}$ 2.5 kHz 40/15.5 ms	$^{13}\text{C}$ 6 kHz 76/12.5 ms	-	32 1.61 s 139 h	596
CANCO	3D 14.1 15	CP H-C, 0.3 ms $^1\text{H}$ : 59 kHz (ramp) $^{13}\text{C}$ : 44 kHz	CP CA-N, 5.5 ms $^{15}\text{N}$ : 11 kHz $^{13}\text{C}$ : 4.7 kHz (ramp)	CP N-CO, 5.5 ms $^{15}\text{N}$ : 11 kHz $^{13}\text{C}$ : 4.7 kHz (ramp)	-	$^{13}\text{CO}$ 40.3 kHz 1612/20 ms	$^{15}\text{N}$ 2.5 kHz 40/15.5 ms	$^{13}\text{C}$ 6 kHz 76/12.5 ms	-	16 2 s 110 h	499
NcoCAB	3D 14.1 15	CP H-N, 1.1 ms $^1\text{H}$ : 76 kHz (ramp) $^{15}\text{N}$ : 60 kHz	CP N-CO, 5 ms $^{15}\text{N}$ : 11 kHz $^{13}\text{C}$ : 4 kHz (ramp)	C-C DARR, 50 ms $^1\text{H}$ RF: 15 kHz	CA-CB DREAM, 4 ms $^{13}\text{C}$ avg RF: 7 kHz tan ramp (54 to 146%)	$^{13}\text{C}$ 50 kHz 1800/18 ms	$^{15}\text{N}$ 2400 Hz 30/12.1 ms	$^{13}\text{C}$ 10 kHz 84/8.3 ms	-	16 1.85 s 87 h	239
CCC	3D 14.1 15	CP H-C, 2 ms $^1\text{H}$ : 76 kHz (ramp) $^{13}\text{C}$ : 60 kHz	C-C DREAM, 5 ms $^{13}\text{C}$ avg RF: 7 kHz tan ramp (56 to 144%)	C-C DARR, 80 ms $^1\text{H}$ RF: 15 kHz	-	$^{13}\text{C}$ 40.3 kHz 1210/15 ms	$^{15}\text{N}$ 13 kHz 140/10.7 ms	$^{13}\text{C}$ 29 kHz 260/8.8 ms	-	2 1.37 s 120 h	704
CONCAB	4D 14.1 15	CP H-C, 3 ms $^1\text{H}$ : 77 kHz (ramp) $^{13}\text{C}$ : 60 kHz	CP CO-N, 6 ms $^{15}\text{N}$ : 11.5 kHz $^{13}\text{C}$ : 4 kHz (ramp)	CP N-CA, 5 ms $^{15}\text{N}$ : 11.5 kHz $^{13}\text{C}$ : 4 kHz (ramp)	CA-CB DREAM, 4 ms $^{13}\text{C}$ avg RF: 7 kHz tan ramp (65 to 135%)	$^{13}\text{CB}$ 50 kHz 1600/16 ms	$^{15}\text{N}$ 2.3 kHz 14/5.6 ms	$^{13}\text{CA}$ 2.5 kHz 34/5.9 ms	$^{13}\text{CO}$ 5.6 kHz 34/5.9 ms	4 1.7 s 162 h	279
CANCOX	4D 14.1 15	CP H-C, 0.9 ms $^1\text{H}$ : 77 kHz (ramp) $^{13}\text{C}$ : 60 kHz	CP CA-N, 7 ms $^{15}\text{N}$ : 11.5 kHz $^{13}\text{C}$ : 4 kHz (ramp)	CP N-CO, 6 ms $^{15}\text{N}$ : 11.5 kHz $^{13}\text{C}$ : 4 kHz (ramp)	CA-CB DARR, 18 ms $^1\text{H}$ RF: 15 kHz	$^{13}\text{C}$ 40.3 kHz 1612/20 ms	$^{15}\text{N}$ 5.6 kHz 27/4.6 ms	$^{13}\text{CA}$ 2.5 kHz 14/5.6 ms	$^{13}\text{CO}$ 2.3 kHz 14/5.6 ms	4 1.78 s 161 h	699



Table S3: Acquisition parameters of MAS NMR spectra for resonance assignment from specifically labeled samples. Spectra were recorded on a 600 MHz Agilent spectrometer in 3.2 mm rotors. Three samples were used, labeled uniformly with  $^{15}\text{N}$  by addition of  $^{15}\text{NH}_4$  to the growth medium, and labeled specifically with the specified amino acids (specified in the table), by addition of these amino acids with  $u\text{-}^{15}\text{N}$ ,  $^{13}\text{C}$  labeling. An exception is the labeling of Leu and Val: here, not the amino acids but the  $^{13}\text{C}$ -labeled acetolactate was added<sup>401</sup>, as a consequence,  $^{13}\text{C}$  is incorporated at the  $\text{C}^\alpha$  position of Val, but not of Leu, and Leu residues are unobservable in all experiments going through  $\text{C}^\alpha$ .

Experiment	Dim. $B_0/T$ MAS	Transfer 1	Transfer 2	Transfer 3	direct dimension spectr. width/kHz points/ $t_{\text{max}}$	ind. dim. 1 spectr. width/kHz points/ $t_{\text{max}}$	ind. dim. 2 d1/s points/ $t_{\text{max}}$	NS time/h	peaks
$u\text{-}^{15}\text{N}$ , GYFR- $^{13}\text{C}$ -labeled sample									
NCACX	3D	CP H-N, 1.1 ms	CP N-CA, 7 ms	C-C DARR, 150 ms	$^{13}\text{C}$	$^{15}\text{N}$	$^{13}\text{C}$	8	310
	14.1	$^1\text{H}$ : 71 kHz (ramp)	$^{15}\text{N}$ : 11 kHz	$^1\text{H}$ RF: 15 kHz	50 kHz	4 kHz	6 kHz	1.9 s	
	15	$^{15}\text{N}$ : 54 kHz	$^{13}\text{C}$ : 4 kHz		1700/17 ms	56/13.7 ms	44/7.1 ms	45 h	
NCOCX	3D	CP H-N, 1.1 ms	CP N-CA, 7 ms	C-C DARR, 150 ms	$^{13}\text{C}$	$^{15}\text{N}$	$^{13}\text{C}$	8	310
	14.1	$^1\text{H}$ : 71 kHz (ramp)	$^{15}\text{N}$ : 11 kHz	$^1\text{H}$ RF: 15 kHz	50 kHz	3.7 kHz	6 kHz	1.9 s	
	15	$^{15}\text{N}$ : 54 kHz	$^{13}\text{C}$ : 4 kHz		1700/17 ms	48/12.7 ms	50/8.1 ms	44 h	
coNCA	2D	CP H-C, 4 ms	CP CO-N, 7 ms	CP N-CA, 7 ms	$^{13}\text{C}$	$^{15}\text{N}$	-	512	12
	14.1	$^1\text{H}$ : 88 kHz (ramp)	$^{15}\text{N}$ : 11 kHz	$^{15}\text{N}$ : 11 kHz	40.3 kHz	2.5 kHz		1.8 s	
	15	$^{13}\text{C}$ : 71 kHz	$^{13}\text{C}$ : 4 kHz	$^{13}\text{C}$ : 4 kHz	1370/17 ms	40/15.6 ms		21 h	
caNCO	2D	CP H-C, 1 ms	CP CA-N, 6 ms	CP N-CO, 6.5 ms	$^{13}\text{C}$	$^{15}\text{N}$	-	128	14
	14.1	$^1\text{H}$ : 88 kHz (ramp)	$^{15}\text{N}$ : 11 kHz	$^{15}\text{N}$ : 11 kHz	50 kHz	2.5 kHz		1.7 s	
	15	$^{13}\text{C}$ : 71 kHz	$^{13}\text{C}$ : 4 kHz	$^{13}\text{C}$ : 4 kHz	1500/15 ms	40/15.6 ms		3.6 h	
$u\text{-}^{15}\text{N}$ , ILV- $^{13}\text{C}$ -labeled sample									
NCACX	3D	CP H-N, 1.1 ms	CP N-CA, 6.5 ms	C-C DARR, 150 ms	$^{13}\text{C}$	$^{15}\text{N}$	$^{13}\text{C}$	24	195
	14.1	$^1\text{H}$ : 71 kHz (ramp)	$^{15}\text{N}$ : 11 kHz	$^1\text{H}$ RF: 15 kHz	50 kHz	3.6 kHz	6 kHz	1.98 s	
	15	$^{15}\text{N}$ : 54 kHz	$^{13}\text{C}$ : 4 kHz		1500/15 ms	30/8 ms	36/5.8 ms	62 h	
NCOCX	3D	CP H-N, 1.1 ms	CP N-CA, 7 ms	C-C DARR, 150 ms	$^{13}\text{C}$	$^{15}\text{N}$	$^{13}\text{C}$	8	237
	14.1	$^1\text{H}$ : 71 kHz (ramp)	$^{15}\text{N}$ : 11 kHz	$^1\text{H}$ RF: 15 kHz	50 kHz	3 kHz	3 kHz	1.7 s	
	15	$^{15}\text{N}$ : 54 kHz	$^{13}\text{C}$ : 4 kHz		1700/17 ms	26/8.3 ms	20/6.3 ms	33.5 h	
coNCA	2D	CP H-C, 4 ms	CP CO-N, 7 ms	CP N-CA, 7 ms	$^{13}\text{C}$	$^{15}\text{N}$	-	8	8
	14.1	$^1\text{H}$ : 88 kHz (ramp)	$^{15}\text{N}$ : 11 kHz	$^{15}\text{N}$ : 11 kHz	40.3 kHz	2.5 kHz		1.8 s	
	15	$^{13}\text{C}$ : 71 kHz	$^{13}\text{C}$ : 4 kHz	$^{13}\text{C}$ : 4 kHz	1370/17 ms	40/15.6 ms		21 h	
CCCDarrdarr	3D	CP H-C, 4 ms	C-C DARR, 120 ms	C-C DARR, 120 ms	$^{13}\text{C}$	$^{13}\text{C}$	$^{13}\text{C}$	2	267
	14.1	$^1\text{H}$ : 88 kHz (ramp)	$^1\text{H}$ RF: 12 kHz	$^1\text{H}$ RF: 12 kHz	40.3 kHz	30 kHz	30 kHz	1.15 s	
	15	$^{13}\text{C}$ : 71 kHz			1210/15 ms	130/4.3 ms	136/4.3 ms	49 h	
$u\text{-}^{15}\text{N}$ , LKP- $^{13}\text{C}$ -labeled sample									
NCACX	3D	CP H-N, 1.1 ms	CP N-CA, 5.5 ms	C-C DARR, 80 ms	$^{13}\text{C}$	$^{15}\text{N}$	$^{13}\text{C}$	8	145
	14.1	$^1\text{H}$ : 71 kHz (ramp)	$^{15}\text{N}$ : 11 kHz	$^1\text{H}$ RF: 15 kHz	50 kHz	3.6 kHz	6 kHz	1.8 s	
	15	$^{15}\text{N}$ : 54 kHz	$^{13}\text{C}$ : 4 kHz		1500/15 ms	56/11 ms	52/8.5 ms	54 h	
NCOCX	3D	CP H-N, 1.1 ms	CP N-CA, 5.5 ms	C-C DARR, 80 ms	$^{13}\text{C}$	$^{15}\text{N}$	$^{13}\text{C}$	8	237
	14.1	$^1\text{H}$ : 71 kHz (ramp)	$^{15}\text{N}$ : 11 kHz	$^1\text{H}$ RF: 15 kHz	50 kHz	4.6 kHz	6 kHz	1.8 s	
	15	$^{15}\text{N}$ : 54 kHz	$^{13}\text{C}$ : 4 kHz		1500/15 ms	56/11.9 ms	46/7.5 ms	43.8 h	
coNCA	2D	CP H-C, 3 ms	CP CO-N, 7 ms	CP N-CA, 7 ms	$^{13}\text{C}$	$^{15}\text{N}$	-	576	25
	14.1	$^1\text{H}$ : 85 kHz (ramp)	$^{15}\text{N}$ : 11 kHz	$^{15}\text{N}$ : 11 kHz	40.3 kHz	3.2 kHz		1.7 s	
	15	$^{13}\text{C}$ : 70 kHz	$^{13}\text{C}$ : 4 kHz	$^{13}\text{C}$ : 4 kHz	1370/17 ms	40/12.2 ms		23.5 h	

Table S4: Acquisition parameters of MAS NMR spectra for resonance assignment using proton-detected experiments on u- $[^2\text{H}, ^{13}\text{C}, ^{15}\text{N}]$ -labeled sample, back-exchanged to 100% in  $\text{H}_2\text{O}$ -buffer. The hcaCBcaNH experiment has not been used for the automatic assignment, but only collected afterwards to confirm assignments manually.

Experiment	Dim. $B_0/T$ MAS	Transfer 1	Transfer 2	Transfer 3	Transfer 4	direct dimension spectr. width/kHz points/ $t_{\text{max}}$	ind. dim. 1 spectr. width/kHz points/ $t_{\text{max}}$	ind. dim. 2 spectr. width/kHz points/ $t_{\text{max}}$	NS d1/s time/h	peaks
hNH	2D 14.1 55	CP H-N, 1 ms $^1\text{H}$ : 95 kHz (ramp) $^{15}\text{N}$ : 40 kHz	CP N-H, 1 ms $^1\text{H}$ : 95 kHz (ramp) $^{15}\text{N}$ : 40 kHz	-	-	$^1\text{H}$ 20.1 kHz 1210/30 ms	$^{15}\text{N}$ 3.04 kHz 112/36.5 ms	-	128 0.95 s 8 h	146
hCONH	3D 14.1 38	CP H-CO, 3 ms $^1\text{H}$ : 95 kHz (ramp) $^{15}\text{N}$ : 40 kHz	CP CO-N, 6 ms $^{15}\text{N}$ : 22 kHz (ramp) $^{13}\text{C}$ : 13 kHz	CP N-H, 1 ms $^1\text{H}$ : 95 kHz (ramp) $^{15}\text{N}$ : 40 kHz	-	$^1\text{H}$ 20.1 kHz 1210/30 ms	$^{15}\text{N}$ 2.8 kHz 46/16 ms	$^{13}\text{CO}$ 3 kHz 50/16.3 ms	128 0.95 s 44 h	304
hCANH	3D 14.1 38	CP H-CO, 3 ms $^1\text{H}$ : 95 kHz (ramp) $^{15}\text{N}$ : 40 kHz	CP CO-N, 6 ms $^{15}\text{N}$ : 22 kHz (ramp) $^{13}\text{C}$ : 13 kHz	CP N-H, 1 ms $^1\text{H}$ : 95 kHz (ramp) $^{15}\text{N}$ : 40 kHz	-	$^1\text{H}$ 20.1 kHz 1210/30 ms	$^{15}\text{N}$ 2.8 kHz 46/16 ms	$^{13}\text{CO}$ 3 kHz 68/9.6 ms	16 1.04 s 64 h	255
hcoCAcoNH	3D 14.1 38	CP H-CO, 3 ms $^1\text{H}$ : 95 kHz (ramp) $^{15}\text{N}$ : 40 kHz	CO-CA out/back INEPT delay 5 ms	CO-N CP $^{15}\text{N}$ : 22 kHz (ramp) $^{13}\text{C}$ : 13 kHz	CP N-H, 1 ms $^1\text{H}$ : 95 kHz (ramp) $^{15}\text{N}$ : 40 kHz	$^1\text{H}$ 20.1 kHz 1210/30 ms	$^{15}\text{N}$ 2.3 kHz 40/17 ms	$^{13}\text{CA}$ 15 kHz 106/7.0 ms	16 1.1 s 92 h	169
hcaCBcaNH	3D 14.1 53	CP H-CA, 4 ms $^1\text{H}$ : 85 kHz (ramp) $^{15}\text{N}$ : 33 kHz	CA-CB out/back INEPT delay 6.5 ms	CA-N CP $^{15}\text{N}$ : 38 kHz (ramp) $^{13}\text{C}$ : 14 kHz	CP N-H, 1 ms $^1\text{H}$ : 82.5 kHz (ramp) $^{15}\text{N}$ : 38 kHz	$^1\text{H}$ 20.1 kHz 1210/30 ms	$^{15}\text{N}$ 2.55 kHz 64/25 ms	$^{13}\text{CA}$ 11.3 kHz 192/8.4 ms	16 1.1 s 58 h	not used in FLYA

Table S5: Acquisition parameters of experiments used for distance restraint measurements.

Experiment	Dim. B <sub>0</sub> /T sample	Through-bond transfer	Through-space transfer	direct dimension spectr. width/kHz points/t <sub>max</sub>	ind. dim. 1 spectr. width/kHz points/t <sub>max</sub>	ind. dim. 2 spectr. width/kHz points/t <sub>max</sub>	NS dl/s time/h
h(N/C)h-RFDR-h(N/C)-H MAS NMR MAS: 55 kHz	3D 14.1 sample 6	CP H-N & H-C, 2.5 ms <sup>1</sup> H: 95 kHz (ramp) <sup>15</sup> N/ <sup>13</sup> C: 41/43 kHz	<sup>1</sup> H RFDR, 8 ms (440 cycles) 100 kHz <sup>1</sup> H pulses	<sup>1</sup> H 12.02 kHz 1200/50 ms	<sup>15</sup> N & <sup>13</sup> C 6.8 kHz 164/24.1 ms	<sup>15</sup> N & <sup>13</sup> C 6.8 kHz 168/24.7 ms	4 0.83 s 118 h
LKP <sup>13</sup> C DARR MAS NMR MAS: 13 kHz	2D 14.1 sample 4	CP H-C, 1.6 ms <sup>1</sup> H: 70 kHz (ramp) <sup>13</sup> C: 55 kHz	<sup>13</sup> C- <sup>13</sup> C DARR, 350 ms <sup>1</sup> H RF: 12 kHz	<sup>13</sup> C 83.3 kHz 2500/15 ms	<sup>13</sup> C 30 kHz 200/6.6 ms	-	192 1.6 s 42 h
H-H-C NOESY solution-state NOESY-HMQC	3D 18.8 sample 7	HMQC type delay 4 ms	<sup>1</sup> H- <sup>1</sup> H NOESY, 400 ms	<sup>1</sup> H 8 kHz 560/70 ms	<sup>1</sup> H 1.5 kHz 32/20 ms	<sup>13</sup> C 3.6 kHz 72/20 ms	12 1.6 s 64 h
H-C-C NOESY solution-state HMQC-NOESY-HMQC	3D 22.3 samples 8-10, see Figure S5E	HMQC type delay 4 ms	<sup>1</sup> H- <sup>1</sup> H NOESY, 350 ms	<sup>1</sup> H 11.4 kHz 571/50 ms	<sup>13</sup> C 4.8 kHz 95/20 ms	<sup>13</sup> C 4.8 kHz 70/14.7 ms	4 1.65 s 60 h

Table S6: Refinement of TET2 structures with *phenix.real\_space\_refine*. Shown are structure statistics obtained when using the three different EM maps (at 4.1, 6 or 8 Å resolution, respectively), and using as starting models different structures, either the crystal structure, or the structures obtained from the NMR and EM approach, that used 4.1, 6 or 8 Å resolution EM maps, respectively.

Model	RMSD w.r.t. PDB 1y0r all atoms	CC <sub>mask</sub>	Ramachandran favored region (%)	All-atom molprobt clashscore	Rotamer outliers (%)
Biological unit derived from X-ray structure (1y0r)	-	0.7836	97.24	6.08	2.21
Refinement of biological unit derived from X-ray structure (1y0r) in cryo-EM map at 4.1 Å, using phenix.real_space_refine (simulated annealing and minimization)	0.739	0.8597	92.71	6.77	11.03
After Steps 1, 2, 3 with 4.1 Å EM map	1.06	0.6668	92.02	161.29	8.56
After Steps 1, 2, 3 at 4.1 Å, refined in cryo- EM map at 4.1 Å, using phenix. real_space_refine	1.087 1.058 1.025	0.74 0.737 0.728	89.74 90.03 90.03	1.71 12.07 3.16	1.03 8.22 26.03
After Steps 1, 2, 3 with 6 Å EM map	2.467	0.418	81.48	33.93	38.24
After Steps 1, 2, 3 with 6 Å EM map refined in cryo-EM map at 6 Å using phenix.real_space_refine (sim. annealing and minimization) (sim. annealing and minimization)	2.157	0.5478	82.62	5.01	0.34
After Steps 1, 2, 3 with 6 Å EM map refined in cryo-EM map at 4.1 Å using phenix.real_space_refine (sim. annealing and minimization)	2.112	0.5553	83.19	5.61	0.34
(After Steps 1, 2, 3 with 8 Å EM map	3.082	0.227	83.52	77.46	44.52
(After Steps 1, 2, 3 with 8 Å EM map refined in cryo-EM map at 8 Å using phenix.real_space_refine (sim. annealing and minimization)	3.111	0.4611	82.62	6.35	0.34
(After Steps 1, 2, 3 with 8 Å EM map refined in cryo-EM map at 4.1 Å using phenix.real_space_refine (sim. annealing and minimization)	3.262	0.4803	82.34	5.9	0