

# **The design of a molecular fixture for cryo-EM structure determination**

Thomas G. Martin<sup>1</sup>, Tanmay A.M. Bharat<sup>1,3</sup>, Andreas C. Joerger<sup>1,4</sup>, Xiaochen Bai<sup>1</sup>, Florian Praetorius<sup>2</sup>, Alan R. Fersht<sup>1</sup>, Hendrik Dietz<sup>2+</sup> and Sjors H.W. Scheres<sup>1+</sup>

<sup>1</sup> MRC Laboratory of Molecular Biology, Francis Crick avenue, Cambridge Biomedical Campus, CB2 0QH, Cambridge, UK

<sup>2</sup> Physik Department, Walter Schottky Institute, Technische Universität München, Am Coulombwall 4a, 85748 Garching near Munich, Germany

<sup>3</sup> Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, United Kingdom

<sup>4</sup> German Cancer Consortium (DKTK), Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe-University, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany

+Correspondence to: dietz@tum.de or scheres@mrc-lmb.cam.ac.uk

## **Abstract**

Despite the recent rapid progress in cryo-electron microscopy, there still exist ample opportunities for improvement in sample preparation. Macromolecular complexes may disassociate or adopt non-random orientations against the extended air-water interface that exists for a short time before the sample is frozen. We designed a hollow support structure using 3D DNA origami in order to protect complexes from the detrimental effects of cryo-EM sample preparation. For a first proof-of-principle, we concentrated on the

transcription factor p53, which binds to specific DNA sequences on double-stranded DNA. The support structures spontaneously form monolayers of pre-oriented particles in a thin film of water, and offer advantages in particle picking and sorting. By controlling the position of the binding sequence on a single helix that spans the hollow support structure, we also sought to control the orientation of individual p53 complexes. Although the latter did not yet yield the desired results, the support structures did provide partial information about the relative orientations of individual p53 complexes. We used this information to calculate a tomographic 3D reconstruction, and refined this structure to a final resolution of approximately 15 Å. This structure settles an ongoing debate about the symmetry of the p53 tetramer bound to DNA.

## Impact statement

As the scope of macromolecular structure determination by electron cryomicroscopy (cryo-EM) is expanding rapidly, it is becoming increasingly clear that many biological complexes are too fragile to withstand the harsh conditions involved in making cryo-EM samples. We describe an original approach to protect proteins from harmful forces during cryo-EM sample preparation by enclosing them inside a three-dimensional support structure that we designed using DNA origami techniques. By binding the transcription cofactor p53 to a specific DNA sequence, and by modifying the position of this sequence in our support structure, we also sought to control the relative orientation of individual p53:DNA complexes.

## Introduction

Cryo-electron microscopy (cryo-EM) structure determination of biological macromolecules is undergoing rapid progress. With the advent of efficient direct electron detectors and the development of powerful algorithms for image processing, numerous structures to near-atomic resolution have been

reported in the past few years{Bai:2015cn, Cheng:2015fo}. In cryo-EM single-particle analysis, solutions of purified protein and/or nucleic acid complexes are applied to a thin sample support, typically an amorphous carbon film with micrometer-sized holes in it that is held in place by a metal grid. Excess liquid is then blotted away with filter paper, and the sample is rapidly plunged in liquid ethane{Adrian:1984dw, Dubochet:1988ks}. This procedure ideally results in the formation of a film of vitreous ice that is only slightly thicker than the macromolecular complex of interest. Keeping the frozen sample at liquid nitrogen temperatures allows its insertion into the high vacuum of a transmission electron microscope, and limits the effects of radiation damage by the electrons that are used for imaging{Taylor:1976bt}. Images taken through the holes of the carbon film ideally contain two-dimensional projections of many, assumedly identical copies of the macromolecular complex of interest, which are often called particles. Projections from different viewing directions can then be combined in a three-dimensional reconstruction of the scattering potential of the molecule{Frank:1981ib}. If the resulting map approaches a resolution of 3 Å it allows building an atomic model of the molecules, from which useful information about their function may be derived.

A major hurdle in single particle analysis is the need to recover the relative viewing angles of the individual particles. This information is lost in the experiment because every particle adopts an uncontrolled orientation in the ice layer. The viewing angles are, therefore, determined *a posteriori* by image processing algorithms that match the experimental projection of every individual particle with projections of a three-dimensional model{Penczek:1992gx}. But, the projection matching procedure is ultimately hampered by radiation damage. Because the electrons that are used for imaging destroy the very structures of interest, see {Glaeser:2016hp} for a recent review, one needs to limit carefully the number of electrons used for imaging. This procedure results in high levels of experimental noise, which in

turn lead to errors in the *a posteriori* determination of the viewing angles. These errors impose severe limitations on the three-dimensional reconstruction, in particular for smaller complexes, as the signal-to-noise ratio in the images decreases with the size of the particles. If one could experimentally control the orientations of each particle in the ice layer, then in principle one could determine structures to higher resolution, and of smaller complexes. This situation is illustrated by samples where the relative orientations of many molecules is set, for example in two-dimensional crystals or helical assemblies of protein molecules. In such cases, near-atomic resolution reconstructions were achieved already decades ago, and from much smaller molecules than currently possible with single-particle analysis{Henderson:1990du, Kuhlbrandt:1994ek, Nogales:1998gb, Gonen:2005em}. Both developments that triggered the recent revolution in attainable resolution of cryo-EM single-particle analysis directly addressed this same hurdle. Better detectors led to lower levels of experimental noise, and better image processing algorithms led to more accurate viewing angles.

Another complication of cryo-EM structure determination lies in the way the sample is prepared, see also {Glaeser:2016fpa, Dobro:2010kd}. The exact physics of cryo-EM sample preparation is poorly described, but several factors may negatively affect its results. Firstly, the blotting process itself may involve strong forces in the sample that destroy fragile protein complexes. Secondly, during the short time between blotting and vitrification, the macromolecules are in a thin liquid film that extends for millimeters to the side, but is only a few hundred Ångstroms thick. Brownian motion will cause the macromolecules to collide with the air-water interface over a thousand times per second{Taylor:2008fy}. Biological macromolecules may unfold when they hit the air-water interface {Glaeser:2016bj}, or they may adsorb to this interface in a non-random manner, for example by presenting their most hydrophobic patch to it. This leads to an uneven distribution of viewing angles in the cryo-EM images, and the corresponding lack of different views

may hamper three-dimensional reconstruction. Despite the numerous successes of single-particle analysis in recent years, preparing suitable cryo-EM samples remains, therefore, a non-trivial task. Often, establishing optimal ice thickness and particle distribution in the ice requires careful optimisation of blotting and freezing conditions by an experienced experimentalist, and fragile complexes have often been observed to fall apart {Thompson:2016tf, Stark:2016bt}.

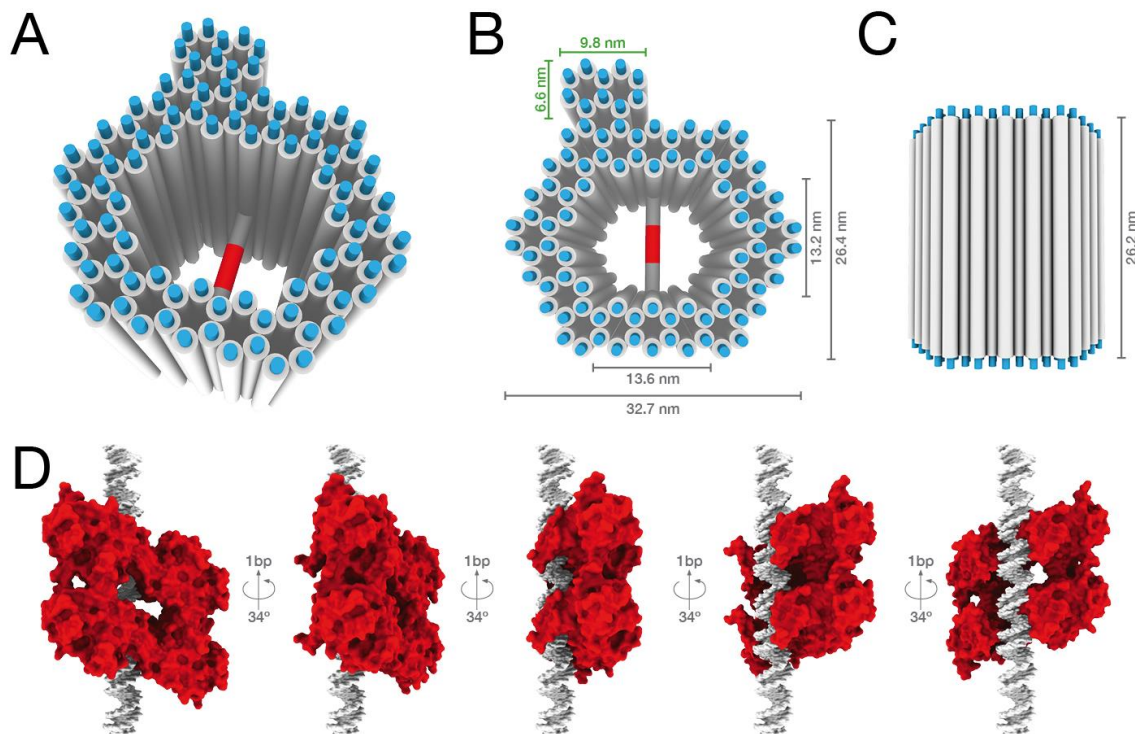
This paper describes an attempt to address both the problems of the unknown viewing angles and the unfavorable conditions of sample preparation from an experimental perspective. To this end, we set out to design a support structure that simultaneously exerts control over the orientations of individual particles, while also controlling the ice thickness and protecting the particles from the air-water interface. We chose to use three-dimensional DNA-origami techniques to design such a support structure. DNA-origami allows flexible and customizable design of three-dimensional structures at the nanometer scale{Douglas:2009dd, Dietz:2009kw, Castro:2011cw}. The use of two-dimensional arrays of DNA-scaffold to bind proteins for cryo-EM sample preparation has been proposed before {Malo:2005fs}. Many proteins naturally interact with nucleic acids like DNA or RNA. In the first instance, to limit the number of technological challenges, we designed a support structure for proteins that naturally bind to double-stranded DNA in a sequence-dependent fashion. This includes many proteins involved in the regulation of transcription, and we chose the transcription factor p53 as a paradigm.

p53 plays a central role in the cell cycle{Lane:2010da}, and is best known for its function in tumor suppression{Joerger:2016fx}. The active form of human p53 is a homotetramer of 4 x 393 amino acids. Its domain organization consists of an intrinsically disordered N-terminal transactivation domain, a proline-rich region, a structured DNA-binding domain (DBD), a tetramerization

domain connected via a flexible linker, and an intrinsically disordered C-terminal regulatory domain{Joerger:2008jk}. The structures of the DBD and the tetramerization domain of human p53 have been solved by x-ray crystallography and NMR{ Canadillas:2006ey, Wang:2007fy, Natan:2011ii, Jeffrey:1995vl, Mittl:1998ht, Clore:1995ts, Lee:1994ka}, as have tetrameric complexes of the DBD with fragments of DNA{Cho:1994wk, Kitayner:2010bl, Chen:2013gq}. The structures of full-length and truncated mutants of human p53 bound to DNA have been analyzed by negative-stain EM and small-angle x-ray diffraction{Tidow:2007ex, Melero:2011gz}, and the DBDs have the same symmetry as found in the crystallographic studies. Other cryo-EM studies on mouse p53 propose a different symmetry{Okorokov:2006fm, Aramayo:2011ic}. As the C-terminus of p53 increases unspecific binding to DNA{Anderson:1997jl}, we chose to use the human truncated p53<sub>1-360</sub> construct (with a molecular weight of approximately 160 kDa for the tetramer) in our experiments. In what follows, we describe the rationale behind our approach, the results obtained with this p53 construct, and the opportunities that this type of support structures offer for more general cryo-EM sample preparation procedures.

## Results

### Support structure design



**Figure 1: Design of the support structure.** (A) Perspective view of the support structure. Each dsDNA helix is shown as a white cylinder. The position of the specific binding sequence on the central dsDNA helix is shown in red; ssDNA overhangs ( $T_{10}$ ) are shown in blue. (B) Top view of the support structure. Inner and outer dimensions of the support structure are shown in grey. The dimensions of the asymmetric feature (flag) are shown in green. (C) Side view of the support structure. (D) Illustration of 5 different settings for the p53-specific binding sequence on the central dsDNA helix. A surface representation of the tetrameric DNA-binding domain of p53 is shown in red. In each of the panels, the p53-binding site is shifted one base upwards from left to right. Because of the helical nature of the dsDNA, this shift also results in a rotation of the p53 complex.

Figure 1 illustrates the design of our proposed DNA-origami support structure. We designed a hollow pillar with a honey-combed motif of 82 parallel dsDNA helices with a height of 26 nm. Two parallel arrays of DNA helices create a central cavity of 13.2nm x 13.6nm and outer dimensions of 26.4nm x 32.7nm. A double stranded (ds) DNA helix with a central p53-

specific binding sequence spans the centre of the hollow space. Ten parallel helices at the periphery, which we will call the flag (Fig 1A, B, top left), make the structure asymmetric so that top and bottom views can be readily distinguished. Different aspects of this design aim to address four main objectives of our support structure.

Firstly, we aim to keep the target protein in the middle of a sufficiently thin ice layer and away from the air-water interface. p53 binds preferentially to the sequence GGACATGTCCGGACATGTCC. By including this sequence in the central dsDNA helix (Fig 1A,B shown in red), our target protein will bind to the approximate centre of our support structure. By designing ssDNA overhangs ( $T_{10}$ ) on both the top and bottom faces of the pillar (Fig 1A-C shown in blue), we aimed to expedite a preferred orientation of the support in the ice layer. The ssDNA overhangs have exposed aromatic bases, which make them more likely to interact with the hydrophobic air-water interface than the sides of the pillar, which is more hydrophilic due to its exposed DNA backbones. This rationale was inspired by initial observations of a preferred orientation in the ice layer of a different DNA-origami object, where similar overhangs were added to prevent end-to-end polymerization of the structures{Bai:2012uv} (Supplementary Figure 1). If both faces of the pillar would indeed interact with the air-water interface at the top and bottom of the sample, then the height of the pillar itself would determine the thickness of the ice layer. By designing a distance of approximately 260 Å between the ssDNA overhangs at the top and the bottom of the pillar, the p53 complex, with an expected size less than 120 Å would then not interact with either of the air-water interfaces. In addition, by anchoring the protein in the middle of the ice layer, differences in defocus height that might lead to loss of resolution in relatively thick ice layers{Jensen:2001ci} can be prevented.

Secondly, by effectively enclosing the target protein inside the support structure, we aim to protect it from blotting forces, aggregation, or other



harmful interactions. It is also conceivable that interactions of the target protein with the holey carbon film, with itself (in the form of aggregation), or with the air-water interface may all influence optimal blotting conditions. Therefore, enclosing the target protein inside a DNA-origami structure that is always the same may also have a beneficial effect on the reproducibility of optimal blotting conditions for different target proteins.

Thirdly, even though we enclose the target protein inside the support, we still aim to minimize the overlap in projected densities of the target protein and the support, as this would hamper alignment of the particle. The central dsDNA helix traverses the middle of the hollow pillar, and the target protein binds to (approximately) the centre of the pillar. Thereby, provided the support structure adopts the intended orientation in the ice, with its open top or bottom faces perpendicular to the electron beam, the projected density from the target protein may be separated from the projected density of the support by excising smaller sub-images (also see Figure 2).

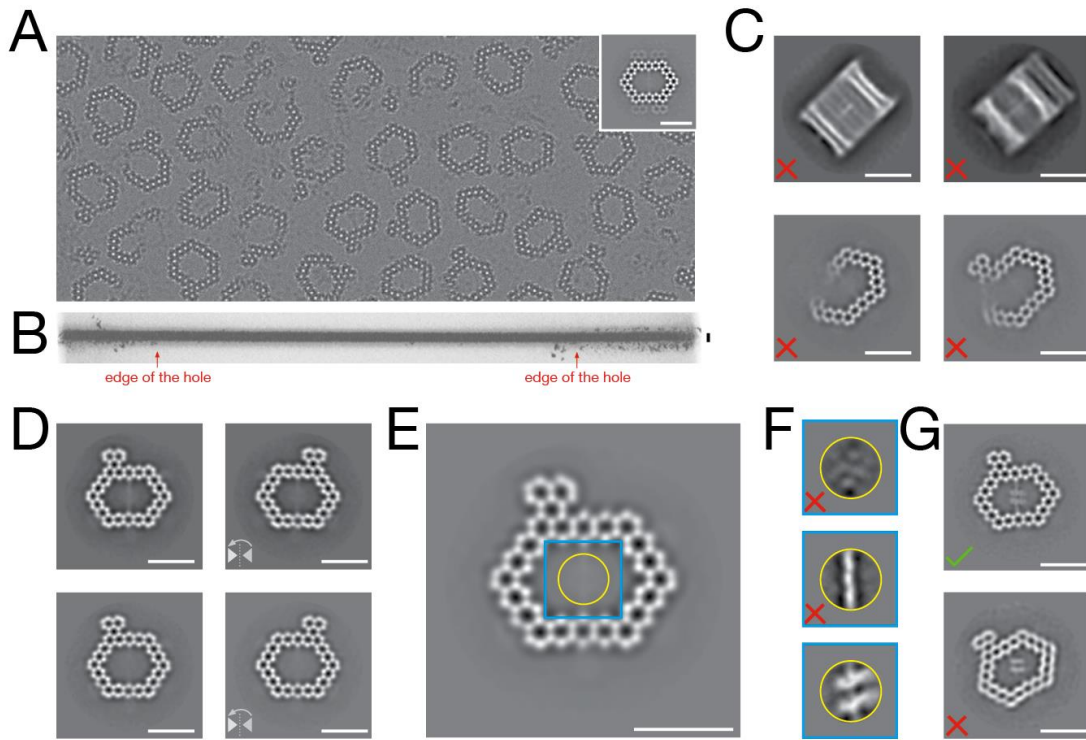
Lastly, we also sought to exert experimental control over the orientation of our target protein, p53 inside the support structure. Because of the helical character of the central dsDNA, we can induce different relative orientations of our target protein with respect to the support structure. By translating the binding sequence one base pair upwards, the p53 complex will rotate  $34^\circ$  along the axis of the dsDNA (Fig 1D). Thereby, by making five different versions of our support structure, shifting the p53-binding sequence one base pair at a time, we can generate different orientations that cover the entire 180 degrees of a tomographic tilt series (Fig 1D). Therefore, we also refer to the central dsDNA helix as the tilt axis. The capability of rotating a single copy of our target protein complex in a controlled manner around the tilt axis turns our design from a passive support into the nano-scale equivalent of a sample holder with a tilting stage. However, as we will discuss in more detail below,

in practice it is difficult to achieve precise control over the orientation of p53 along the tilt axis.

## Synthesis and imaging of the support structure

The designed DNA-origami structure was synthesized and purified with a yield between 50% and 90% using standard procedures (see Materials and Methods). We first used the support structure alone, i.e. without the target protein, to determine suitable freezing conditions for cryo-EM grid preparation. We found conditions where the support structures adopt a pseudo 2D-crystalline arrangement in large areas of the grid, where most structures adopt the intended top-bottom orientation in the ice layer (Fig 2A). Using cryo-electron tomography, we confirmed that the ice layer is indeed as thick as the designed DNA-origami support structures (Fig 2B, Supplementary Movie 1). Interestingly, in areas where the ice appeared thinner than the designed height, we observed no support structures. In areas where the ice appeared thicker, we also observed support structures in side-view orientations (Supplementary Figure 2)

Subsequently, we made cryo-EM grids of the support structure together with p53. We made five different samples, each of which with a different position of the p53 binding sequence on the tilt axis. As hypothesized, we were able to use the same blotting conditions as for the empty support structures, although the grids exhibited fewer areas with the optimal, near-crystalline arrangements of the support structures. Nevertheless, we could use the observation that support structures preferentially adopt top-bottom views in regions with the desired ice thickness to select suitable areas for data acquisition from relatively low-magnification overviews in the microscope (Supplementary Figure 2). Data acquisition was performed for each tilt axis setting separately on an FEI Titan Krios microscope at 300 kV with a K2 Summit detector (Table 1).



**Figure 2: Image processing strategy.** All scale bars are 20nm. (A) Part of a typical micrograph. A template for automated particle picking is shown in the top right corner. (B) Tomographic side view of a hole showing a monolayer of support structures. Near the edge of the hole (indicated with an arrow) the ice gets slightly thicker. (C) Examples of 2D classes that are discarded. (D) Examples of 2D classes of intact structures with the flag in the top-left or top-right. A mirror operation is applied to images with the flag on the top-right. (E) The average of all intact particles with the correct orientation in ice, including the mirrored particles, is used as a template for their alignment. (F) Sub-images (cyan) with a width and height of 20nm are extracted from the aligned particles, and submitted to 2D classification with a prior on the in-plane rotation. A circular mask that is applied during this process is shown in yellow. Three types of particles are distinguished: supports without tilt axis (top), supports with tilt axis but without a density for p53 (middle) and supports with tilt axis and p53 density (bottom). (G) Illustration of the final selection of particles, where the angle from a re-alignment of the p53-only sub-image is compared with the angle of the entire support structure, and particles with large differences in these angles (e.g. bottom panel) are discarded.

309

	-2bp	-1bp	0bp	+1bp	+2bp	Total
<b>Micrographs</b>	604	415	582	576	385	2562
<b>Auto picked particles</b>	69214	51830	61528	65942	24400	272914
<b>Intact top view supports (mirrored)</b>	36727 (19672)	40824 (26886)	41131 (19122)	39303 (19526)	19302 (8988)	177287 (94194)
<b>Empty supports</b>	11018	15916	18452	14424	6319	66129
<b>Supports with only DNA</b>	15439	17082	16587	15665	7412	72185
<b>Supports with p53-like density</b>	10270	7826	6092	9214	5571	38973
<b>p53 in expected orientation</b>	5821	4831	3373	5169	3504	22698
<b>p53 in 2D classes used for initial-model calculation</b>	2363	2197	1375	1795	1541	9271

310 **Table 1:** Data acquisition statistics for each tilt axis setting.

311

## 312 2D image pre-processing

313

314 The incorporation of the target protein inside a larger support structure not  
 315 only provides experimental information about its orientation, it also facilitates  
 316 the selection of individual particles from the micrographs. We use the  
 317 template-based particle selection procedure in RELION for this{Scheres:2015it}.

318 By using a 2D template structure that corresponds to the top view of our  
 319 support structure (Fig 2A, inset), we automatically selected 272,914 particles  
 320 from the five different experiments, and combined all these particles into a  
 321 single data set.

322

323 To remove inadvertently picked side views, broken support structures or other  
 324 false positives (Fig 2C) from the data set, we selected 177,287 particles in  
 325 three rounds of reference-free 2D classification. As the support structure may  
 326 adopt either a top-up or a bottom-up orientation in the ice, we applied a  
 327 mirror operation to the 53% of these particles that contributed to 2D class  
 328 averages with a flag on the top right side of the support structure (Fig 2D,  
 329 Table 1). The resulting set of intact support structures was aligned to a  
 330 common reference (Fig 2E), to allow extraction of the central part of the  
 331 individual particle images in a smaller box. An additional round of reference-

free 2D classification was used to discard particles for which no obvious p53 density was visible (Fig 2F). In this calculation, we used a prior probability, or prior in short, on the in-plane rotations. In the empirical Bayesian approach to image processing in RELION{Scheres:2012bs}, prior probabilities on orientational parameters are expressed as Gaussian functions centered on the expected value for that orientation, and with a standard deviation that expresses the uncertainty in that expectation. In this case, we centered the prior on the in-plane rotations from the initial 2D alignment of the entire support structure and used a standard deviation of 15° to allow for errors in those assignments. At this point 36,837 particle images were selected. A final round of particle selection was based on identifying particles that rotated too much in the p53-only realignment with respect to the support structure (Fig 2G), and led to a final data set of 22,698 sub-images containing p53.

### 3D reconstruction

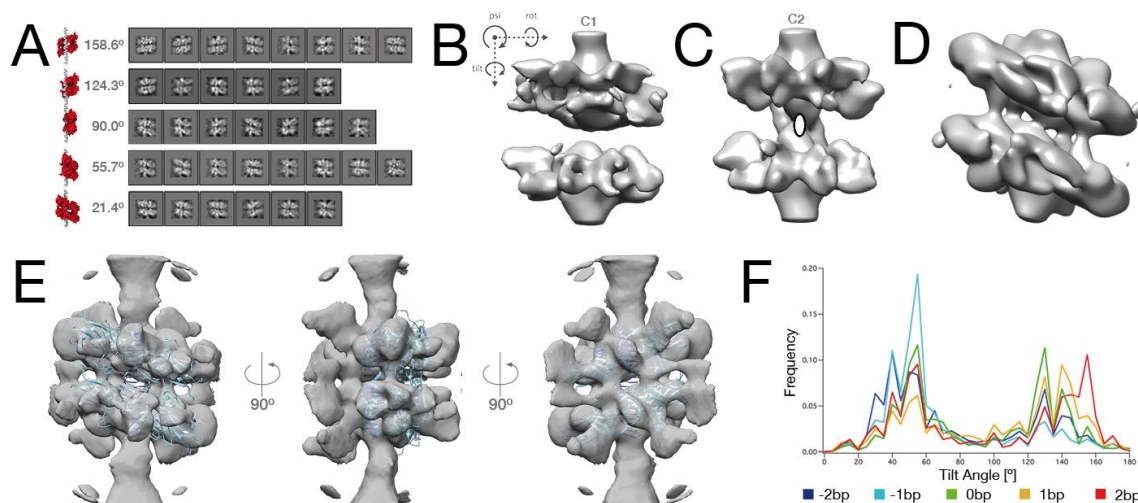
The 2D alignment parameters of the support structures, combined with the known setting of the tilt axis, provide information about the five parameters that define the orientation of every p53 complex. Using only 2D alignment parameters, one can calculate, therefore, a tomographic 3D reconstruction of p53. The two in-plane translations of the support structure, combined with the known offset of the binding sequence on the tilt axis, can be used to centre the target protein. In addition, three Euler angles describe the relative orientations of all p53 particles with respect to a common three-dimensional frame of reference. By using a reference with the dsDNA tilt axis along the Y axis, the first Euler angle ("rot" in RELION) is expected to be around 0°, as the tilt axis is designed to run perpendicular to the top-bottom axis of the support structure. The second Euler angle ("tilt") is determined by the tilt-axis setting of the experiment. By defining the tilt angle for our central tilt axis position (0bp) as 90°, the one or two base-pair shifted positions of the binding sequence on either side have expected tilt angles of 21.4° (-2bp),

55.7° (-1bp), 124.3° (+1bp) and 158.6° (+2bp). Finally, the third Euler angle ("psi") is directly available from the in-plane rotation determined in the 2D alignment of the support structure.

In order to calculate the initial tomographic p53 reconstruction, we first split the selected 22,698 sub-images by their original tilt axis settings, and performed five separate 2D classifications using a prior on the in-plane rotations of 15°. From these five runs, we selected a total of 35 2D class averages with the best protein-like features (Fig 3A). We then used the expected Euler angles as defined above to perform a tomographic reconstruction, directly from the 2D class averages. A preliminary map without symmetry (Fig 3B) indicated the presence of C2 symmetry, which was subsequently imposed (Fig 3C). Apart from the 2-fold rotational symmetry, the reconstructed map also showed local translational pseudo-symmetry along the dsDNA axis.

The observations that particles from different tilt axis settings gave rise to similar 2D class averages, and that different 2D class averages were observed within a single tilt axis data set, indicate that the tilt axis is probably more flexible than anticipated. In principle, the statistical framework of RELION is well-suited to model deviations from the expected orientations of each individual particle through the use of Gaussian priors on the translations and the Euler angles. The standard deviation of the Gaussian priors can be used to tune the amount of expected deviations. Larger deviations from the expected orientations, e.g. because the attachment of the target protein to the DNA-origami structure is more flexible than anticipated, will lead to less-informative priors. To some extent, deviations from the anticipated tilt angle are actually beneficial to the reconstruction process, as they will lead to a more uniform angular sampling along the tomographic tilt axis.

We first performed, therefore, a rotational and translational re-alignment of the 35 selected 2D class averages, where we used a prior with a standard deviation of  $10^\circ$  on rot and psi, but tilt was left free. The resulting reconstruction (Fig 3D) was then used as an initial model in a set of 3D refinements, where we used the 9,271 p53 particles that were assigned to the selected 2D classes. Without using rotational priors on any of the Euler angles, or when using only a prior on rot or psi, the reconstructions looked worse (e.g. density for the dsDNA disappeared) than when using a  $5^\circ$  prior on both rot and psi. As expected from the observation that our tilt axis is more flexible than anticipated, imposing a prior on tilt made the reconstruction worse (Supplementary Figure 3). When using progressively less informative, i.e. broader, priors on rot and psi, the reconstructions also became worse (Supplementary Figure 4), whereas priors, with standard deviations less than  $5^\circ$  are too narrow for the angular sampling rate employed in the refinement. Consequently, our best refinement had a standard deviation of  $5^\circ$  for the priors on psi and rot, and left the tilt angle unrestrained. The resulting map, at an estimated resolution of approximately 15 Å, shows the same domain architecture as the known crystal structure of the DNA-binding domain of p53 bound to dsDNA{Chen:2013gq} and shows additional densities in both the front and the back of the complex (Fig 3E). The histogram of the refined tilt angles confirms that the tilt axis is more flexible than anticipated. Some information in the tilt angle is maintained, but angles around  $90^\circ$  are somehow disfavored (Fig 3F).



**Figure 3: Initial model generation and final maps.** (A) Class averages used for the initial model generation sorted by tilt axis setting. An illustration of the designed orientation is shown on the left and the applied tilt angle is shown in front of the class averages. (B) Initial model reconstructed with the expected tilt angles in C1. The angles for rot and psi are all set to 0°. In RELION, the first rotation of the 3D reference object (rot) is around the Z-axis, which comes out of the XY-plane of the figure; the second rotation (tilt) is around the new Y-axis, and the third rotation (psi) is around the new Z-axis. The inset shows a simplified explanation of these rotations from the point of view of the experimental particles. In that case, psi is the in-plane rotation, tilt is the rotation around the central DNA axis, and rot describes out-of-plane rocking. (C) The same model as in B with C2 symmetry imposed. The 2-fold symmetry axis is along the Z-axis, and is indicated with an oval. (D) Map after re-alignment of the class averages with 15° priors on the rot and psi angles and unrestricted tilt angles. (E) Different views of the final map generated from 9,271 particles. The estimated resolution is around 15 Å. PDB model 4HJE30 of the DBD in blue. (F) Histogram of the refined tilt angles for each of the tilt axis settings.

## Discussion

We present a novel approach to sample preparation for cryo-EM structure determination of biological macromolecules. Using 3D DNA-origami, we designed a support structure with a defined size and shape that binds



specifically to a target protein of interest. The resulting structure is the first artificial scaffold that exerts experimental control over the orientations of individual protein molecules, and protects them from aggregation or harmful interactions with the air-water interface. In addition, the support structure may facilitate the optimisation of freezing conditions, and aid in the selection of suitable ice thickness for data acquisition. Nevertheless, as discussed below, this work does not yet represent a ready-to-use solution for high-resolution structure determination of a wide range of different target proteins, but should rather be considered as a proof-of-principle towards achieving this ambitious goal.

By choosing a target protein that naturally binds to dsDNA in a sequence-specific manner, our support structure was designed to act as a molecular fixture that exerts experimental control over the orientations of individual protein complexes. By using five different positions of the p53 binding sequence on a central dsDNA helix in our structure, our structure intended to provide five different views of a p53 tetramer bound to the support structure. Although the prior information about the five different orientations could indeed be used successfully to calculate an initial tomographic reconstruction of the p53 tetramer bound to DNA, further refinement of the orientations of 2D class averages or individual particle images revealed a distribution in tilt angles that is much broader than one would expect from a rigid dsDNA helix, and somehow seems to disfavor tilt angles around 90°. Because the complete structure of the p53 tetramer remains unknown, it could be that the support structure is too small to accommodate p53 in this orientation. Analysis of multiple 2D class average images from our data (Supplementary Movie 2) reveals that the support structures are also not as rigid as one might need to exert precise orientation control. Moreover, despite previous attempts at optimizing the incorporation of the central dsDNA helix in the support structure {Martin:2014wz}, it could be that our current design (Supplementary Figure 5-6) still leads to incorrect attachments in a subset of the structures.

Nevertheless, despite the lack of complete control over the tilt angle, the support structure still provided information about the centre of the particle, its in-plane rotation (the  $\psi$  angle) and its out-of-plane rocking (the  $\theta$  angle), and this information could be used in statistical priors to improve the 3D reconstruction. The intended design of our support structure as a molecular fixture that provides experimental information about the orientations of individual p53 complexes was, therefore, at least partially successful.

Our experiments also revealed several challenges that will need to be overcome in future studies. First of all, the final number of 9,271 selected p53 particle images represents a low yield from 2,562 selected micrographs. Approximately 42% of our support structures did not incorporate the tilt axis, and 77% of the remaining support structures did not bind to p53. The latter was surprising, as at a concentration of 1mM p53 tetramers and 150nM of support structures, and with a dissociation constant of approximately 50nM for dsDNA, we had expected almost all support structures to contain p53. Possibly, steric hindrance from the support, or strain in the central dsDNA affected the binding of p53. A further reduction in particle number came from our image selection procedures: 38% of the particles that contained density for p53 had it bound in an unrealistic orientation in respect to the support structure, and from these only 41% contributed to the selected 2D classes of the sub-images that showed the best protein-like features. Secondly, although the low particle yield may have hampered reaching higher resolution, the selected particles still represented more than 18,000 asymmetric units, which should probably have yielded a higher resolution reconstruction, *cf.* ref {Fernandez:2013fe}. One possibility could be that the central cavity of the support structure was too small for the defocus used. Delocalization in the images due to defocussing may have led to a superposition of the signal from the support structure and the target protein, which may have limited the alignment of the individual target proteins, and therefore the resolution of the

final reconstruction. This problem could in principle be circumvented by in-focus imaging through phase-plate technology {Danev:2016hm}. In addition, future designs of larger, more stable support structures with a more accessible protein binding site, and with better incorporation of the tilt axis may improve both particle yield and resolution. **In order to create these larger fixtures with low defect rates, new assembly and purification strategies will need to be considered and adopted, see for example** {Gerling:2015ix, Shaw:2015iy}.

Nevertheless, at a resolution of approximately 15 Å, the resulting reconstruction provides useful information about how p53 binds to dsDNA. The observation that our reconstruction shows C2 symmetry with an additional 2-fold translational component along the DNA axis, which is similar to the crystal structure of the tetrameric core of p53 bound to dsDNA, indicates that p53 does not bind to DNA in a previously proposed arrangement with D2 symmetry{Okorokov:2006fm, Aramayo:2011ic }. Instead, the presence of extra density in the front and back of the complex (Fig 3E) provides support for an alternative model where the p53 tetramerization domain binds to the opposite side of the dsDNA helix from the DNA-binding domain{Tidow:2007ex}.

The design of our support structure also contains several useful features that may inspire future experiments. By enclosing the target protein in a hollow structure, it is protected from interactions with copies of itself or with the hydrophobic air-water interface during cryo-EM sample preparation. This may prevent the proteins from aggregation, or from unfolding or adopting preferred orientations against this air-water interface. Also, it might be that the optimisation of freezing conditions will depend more on the support structure than on the nature of the protein inside, so that the optimal conditions would differ less for different proteins compared to current methods. We indeed used similar conditions for the samples with only

534 support structures and the mixture with p53, although the mixture showed  
535 fewer regions of near-crystalline arrangements of the support structures. We  
536 used approximately 6 times more p53 tetramers than support structure in our  
537 mixture, and did not attempt to remove unbound p53 tetramers through  
538 additional purification steps. It could therefore be that interactions of  
539 unbound p53 tetramers with the air-water interface still had an effect on the  
540 freezing conditions. In addition, the observation that the support structures  
541 adopt monolayers in an ice layer with the intended thickness is potentially  
542 useful. One might for example try to add support structures to standard cryo-  
543 EM samples with the idea of using them as 'spacers'{Glaeser:2016fp} to  
544 control ice thickness. Admittedly, it remains difficult to conclude whether the  
545 support structures maintain the desired ice thickness over a given area, or  
546 whether the monolayers just happen to form in those areas where the ice  
547 thickness is ideal. But already, because one can easily select areas with  
548 monolayers of support structures from low-magnification overview images in  
549 the microscope, the support structures do facilitate the data acquisition  
550 process. The interactions of our support structure with the air-water interface  
551 may also have a beneficial effect of the local particle concentration. For a 300  
552 Å thick layer of a 150 nM solution, one would expect only a single support  
553 structure in every micrograph. Our observation of approximately 100  
554 structures per micrograph suggests that interactions of the support structure  
555 with the air-water interface, possibly both at the top and the bottom, may  
556 lead to a strong local enrichment in concentration. For proteins, a similar  
557 enrichment has been attributed to a sticky layer of denatured protein at the  
558 air-water interface {Taylor:2008fy}. Because our samples without protein show  
559 a similar, or even stronger, enrichment in DNA support structures, unbound  
560 p53 tetramers are probably not required for the enrichment. It is difficult to  
561 assess whether the interactions with the air-water interface lead to unfolding  
562 of the support structures themselves. We do observe partial structures (Figure  
563 2A,C), but these could also be explained by folding defects. It therefore

remains unclear whether the presence of denatured DNA material at the air-water interface is important for the enrichment effect.

The experiments described here rely on the specific binding of a target protein that naturally binds to dsDNA in a sequence-specific manner. Although several transcription factors are known to do so, the design of a more broadly applicable support structure would be desirable. It would be relatively straightforward to include a DNA mismatch in the central axis, which could include target proteins from DNA mismatch repair pathways. Alternatively, one could construct a central axis with a bubble, or from DNA/RNA hybrids to include a range of different target proteins related to replication, transcription or DNA repair. However, in order to design a support that could bind to an arbitrary target protein, one would probably need to combine chemical modifications of some of the DNA-staples within the support with a specific tag on the target protein. For example, one could use commercially available biotin-labeled DNA on the support structure to bind a counter part to a protein tag attached to a monovalent streptavidin{Fairhead:2014eu, Veggiani:2014hj}. For the study of membrane proteins, one might even consider designs where nano-discs {TimothyHBayburt:2002jn} are attached to DNA-origami scaffolds, or where scaffolds interact directly with patches of membranes {Kocabey:2015da}. Probably, such more generally applicable designs would retain even less control over the orientation of the target protein than the design described in this paper. Still, the scaffold structure could maintain some of its other useful features, such as facilitating particle selection, preventing aggregation and keeping proteins away from harmful interfaces. Apart from the beneficial effects on cryo-EM sample preparation, large support structures that specifically bind a target protein of interest might even play a role in protein purification. Provided the binding of the target protein would be tight enough, their large molecular weight (~5 megadalton) would allow relatively straightforward separation of the target protein from smaller contaminants

through the use of for example gel filtration of sucrose gradients. Alternatively one could try to fix the DNA supports directly on the grid surface and use the grids themselves for on-grid affinity purification in a similar manner as was done with NTA-modified lipid monolayers {Kelly:2008kx, Kelly:2010dm} or antibodies attached to carbon films {Yu:2014kc, Yu:2016eq}.

In summary, this work provides an original approach to exert experimental control over the orientations of individual protein complexes and protect them from harmful forces during cryo-EM sample preparation. As the field of cryo-EM structure determination keeps growing, the physics involved in its sample preparation will become clearer and new concepts in sample preparation will continue to emerge. Our approach provides ample possibilities for further developments, and may contribute, thereby, to tackling some of the outstanding challenges in this rapidly changing technique.

## Acknowledgements

We are grateful to Christos Savva, Shaoxia Chen, Toby Darling and Jake Grimmett for technical support, and Miriana Petrovich for protein purification. This work was funded by EMBO (postdoctoral fellowship ALTF 1229-2013 to TGM; advanced fellowship aALTF 778-2015 to TAMB) and the European Commission (Marie Skłodowska-Curie postdoctoral fellowships to TGM and XCB); the project was further supported by a European Research Council Starting Grant to H.D. (GA no.256270) and by the Deutsche Forschungsgemeinschaft through grants provided within the Sonderforschungsbereich SFB863, the Center for Integrated Protein Science Munich, and the Nano Initiative Munich; and the UK Medical Research Council (MC\_UP\_A024\_1010 to ARF and MC\_UP\_A025\_1013 to SHWS).

The p53 reconstruction was submitted to the Electron Microscopy Data Bank (EMDB accession number XXXX). [to be completed upon acceptance]

## Materials and Methods

### Origami Design, synthesis and purification

DNA-origami design was performed in caDNAo v0.2{Douglas:2009ix}. The scaffold DNA (the 7,560 nucleotide long version of the the M13mp18 phage genome) was prepared as described previously{Douglas:2007ww}. DNA staple oligonucleotides, prepared by solid-phase chemical synthesis, were ordered from Eurofins MWG. The DNA oligos for the tilt axis were ordered to HPLC-grade purity, all other oligos were ordered to high-purity, salt-free grade. The supports containing different tilt-axis settings were synthesized individually in one-pot mixtures containing 50nM scaffold DNA, 75nM tilt axis DNA, and 200nM staple DNA in a 10mM Tris buffer at pH 7.6 with 1mM EDTA and 20mM MgCl<sub>2</sub>. The mixture was incubated at 65°C for 15 min, then annealed from 60°C to 45°C over the course of 8h, and stored at 25°C. Purification from excess staple oligos was performed in 5 rounds of molecular weight cutoff filtration using 100 kDa Amicon filters (Millipore) in a buffer containing 20mM Tris base and 5mM MgCl<sub>2</sub>. The final concentration of the support structures was adjusted to ~250nM.

### P53 expression and purification

Production and purification of truncated p53 variant (residues 1-360) lacking the C-terminal regulatory domain followed published protocols{Veprintsev:2006jo, Natan:2011ii}. Briefly, the proteins were produced in *Escherichia coli* BL21(DE3) as a fusion protein with N-terminal 6×His-tag, *Bacillus stearothermophilus* lipoyl domain, and tobacco etch virus protease

cleavage site. They were then purified using standard His-tag purification protocols, followed by tobacco etch virus protease cleavage, heparin affinity chromatography, and a final gel filtration step on a Superdex 200 16/60 preparative gel filtration column (GE Healthcare) in 300 mM NaCl, 20 mM Tris, pH 7.5, and 5 mM DTT. Protein samples were flash frozen in liquid nitrogen and stored at -80 °C. The variant contained four stabilizing mutations, M133L/V203A/N239Y/N268D{Natan:2011ii, Nikolova:1998ti, Joerger:2004ky}, in the DNA-binding domain.

The purified p53 sample was mixed with the purified DNA origami sample to final concentrations of ~150nM DNA origami, and ~4μM p53 (monomer) in a 20mM Tris buffer containing 1.5μM DTT, 45mM NaCl and 4mM MgCl<sub>2</sub>. The mixture was incubated for 20 minutes at 4 °C, prior to preparing cryo-EM grids.

#### Electron microscopy

Cryo-EM grids for the support structures alone or for the support structures with p53 and the five different settings of the tilt axis were prepared separately using similar procedures. Aliquots of 3μl of sample were incubated for 10 seconds on glow-discharged Quantifoil grids, blotted for 2 s, and plunge-frozen in liquid ethane using a Vitrobot Mark 3 (FEI Company). Grids were transferred to an FEI Titan Krios microscope that was operated at 300 kV. Images were recorded on a K2 detector using a Gatan energy filter (with a slit width of 20 eV). Movies with a total electron dose of ~38 e<sup>-</sup>/Å<sup>2</sup> were recorded in super-resolution mode, and subsequently downsampled to a final pixel size of 1.76 Å in RELION. The defocus was varied between 1 and 5 μm. Cryo-electron tomography data was collected using procedures described previously {Bharat:2015fc}.

Electron micrographs were manually evaluated for astigmatism and drift, and 2,562 micrographs were selected for further analysis. Beam-induced motion



685 correction was performed in UCSF MOTIONCORR{Li:2013gt}; estimation of  
686 contrast transfer function (CTF) parameter was performed in  
687 Gctf{Zhang:2016fy}; and all subsequent image processing operations were  
688 performed in RELION-1.4 {Scheres:2012bs}.

689

690

## 691 **References**

692

693 {papers2\_bibliography}