



## OPEN ACCESS

## EDITED BY

Tao Liu,  
University of New South Wales, Australia

## REVIEWED BY

Bruce A. Shapiro,  
National Cancer Institute at Frederick  
(NIH), United States  
Takeru Hayashi,  
Institute of Microbial Chemistry (IMC),  
Japan

## \*CORRESPONDENCE

Yusuke Suenaga  
✉ysuenaga@chiba-cc.jp  
Tatsuhito Matsuo  
✉matsuo.tatsuhito@qst.go.jp

RECEIVED 28 April 2023

ACCEPTED 03 November 2023

PUBLISHED 23 November 2023

## CITATION

Yamamoto S, Kono F, Nakatani K, Hirose M, Horii K, Hippo Y, Tamada T, Suenaga Y and Matsuo T (2023) Structural characterization of human *de novo* protein NCYM and its complex with a newly identified DNA aptamer using atomic force microscopy and small-angle X-ray scattering. *Front. Oncol.* 13:1213678. doi: 10.3389/fonc.2023.1213678

## COPYRIGHT

© 2023 Yamamoto, Kono, Nakatani, Hirose, Horii, Hippo, Tamada, Suenaga and Matsuo. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Structural characterization of human *de novo* protein NCYM and its complex with a newly identified DNA aptamer using atomic force microscopy and small-angle X-ray scattering

Seigi Yamamoto<sup>1</sup>, Fumiaki Kono<sup>2</sup>, Kazuma Nakatani<sup>1,3,4,5</sup>, Miwako Hirose<sup>6</sup>, Katsunori Horii<sup>6</sup>, Yoshitaka Hippo<sup>1,3,7</sup>, Taro Tamada<sup>2,8</sup>, Yusuke Suenaga<sup>1\*</sup> and Tatsuhito Matsuo<sup>2\*</sup>

<sup>1</sup>Laboratory of Evolutionary Oncology, Chiba Cancer Center Research Institute, Chiba, Japan,

<sup>2</sup>Institute for Quantum Life Science, National Institutes for Quantum Science and Technology, Chiba, Japan, <sup>3</sup>Graduate School of Medical and Pharmaceutical Sciences, Chiba University, Chiba, Japan, <sup>4</sup>Innovative Medicine CHIBA Doctoral WISE Program, Chiba University, Chiba, Japan,

<sup>5</sup>All Directional Innovation Creator Ph.D. Project, Chiba University, Chiba, Japan, <sup>6</sup>Digital Healthcare Business Development Office, NEC Solution Innovators, Ltd., Tokyo, Japan, <sup>7</sup>Laboratory of Precision Tumor Model Systems, Chiba Cancer Center Research Institute, Chiba, Japan, <sup>8</sup>Graduate School of Science, Chiba University, Chiba, Japan

NCYM, a Homininae-specific oncoprotein, is the first *de novo* gene product experimentally shown to have oncogenic functions. NCYM stabilizes MYCN and  $\beta$ -catenin via direct binding and inhibition of GSK3 $\beta$  and promotes cancer progression in various tumors. Thus, the identification of compounds that binds to NCYM and structural characterization of the complex of such compounds with NCYM are required to deepen our understanding of the molecular mechanism of NCYM function and eventually to develop anticancer drugs against NCYM. In this study, the DNA aptamer that specifically binds to NCYM and enhances interaction between NCYM and GSK3 $\beta$  were identified for the first time using systematic evolution of ligands by exponential enrichment (SELEX). The structural properties of the complex of the aptamer and NCYM were investigated using atomic force microscopy (AFM) in combination with truncation and mutation of DNA sequence, pointing to the regions on the aptamer required for NCYM binding. Further analysis was carried out by small-angle X-ray scattering (SAXS). Structural modeling based on SAXS data revealed that when isolated, NCYM shows high flexibility, though not as a random coil, while the DNA aptamer exists as a dimer in solution. In the complex state, models in which NCYM was bound to a region close to an edge of the aptamer reproduced the SAXS data. Therefore, using a combination of SELEX, AFM, and SAXS, the present study revealed the structural properties of NCYM in its functionally active form, thus providing useful information for the possible future design of novel anti-cancer drugs targeting NCYM.

## KEYWORDS

*de novo* protein, NCYM, DNA aptamer, solution structure, AFM, SAXS

## 1 Introduction

NCYM, a *cis*-antisense gene of MYCN, encodes a Homininae-specific oncoprotein (1, 2). In human neuroblastomas, NCYM is always co-amplified with MYCN, and its expression level is associated with poor prognosis (1). NCYM stabilizes MYCN via inhibition of GSK3 $\beta$ , whereas MYCN stimulates both MYCN and NCYM transcription (1). This feedback loop contributes to the maintenance of high levels of both MYCN and NCYM expressions in MYCN-amplified neuroblastomas (1, 2). NCYM enhances the metastasis of neuroblastomas (1) possibly via inhibition of apoptotic cell death (1, 3, 4) and/or regulation of stemness (5, 6). Furthermore, NCYM has been shown to be associated with progression of adult cancers (2, 7). Therefore, NCYM is a promising target protein for anti-cancer therapy. However, the difficulty in determining its structure hinders drug design (8).

NCYM is a newly evolved coding gene that originated from MYCN promoter region during the evolution of the Homininae (1, 2). New genes originating from non-genic regions are known as *de novo* gene birth (9–11), and NCYM is the first *de novo* gene product experimentally shown to have oncogenic functions. Owing to their *de novo* emergence, *de novo* proteins show no homology to known genes and do not have any domains or motifs. The amino acid sequence of *de novo* proteins is similar to a random sequence (12), although a recent report identified the difference between *de novo* proteins and unevolved random-sequence counterparts in that *de novo* proteins exhibit moderately higher solubility in cells (13). Four *de novo* proteins have been structurally characterized to date: Bsc4 (14), NCYM (8), Goddard (15), and AFGP8 (16); however, mainly because of their highly disordered nature, none of the complete structures have been determined. Upon binding to the ice surface, the local structure of the antifreeze glycoprotein AFGP8 make a transition from a disordered to an ordered state (16), indicating the possibility of significant ordering of *de novo* proteins via complex formation with binding partners. Consistent with an earlier prediction that NCYM binds to DNA (17), we have previously found that benzonase treatment significantly improves the solubility of NCYM (8). These observations led us to identify DNA aptamers that bind specifically to NCYM and to consider that analysis of the complex of NCYM and DNA aptamers may contribute to the characterization of the structural dynamics of NCYM.

Here, three types of DNA aptamers were identified by systematic evolution of ligands by exponential enrichment (SELEX) and their interactions with NCYM were characterized by atomic force microscopy (AFM). Moreover, the structure of the NCYM-DNA complex of a representative DNA aptamer (named “No. 1”), which enhances interaction between NCYM and GSK3 $\beta$ , was analyzed using small-angle X-ray scattering (SAXS).

## 2 Materials and methods

### 2.1 Aptamer selection procedure

SELEX was performed as previously reported with some modifications (18, 19). Dynabeads<sup>TM</sup> MyOne<sup>TM</sup> Carboxylic Acid (CA) magnetic beads (Invitrogen, Waltham, MA) were used for

NCYM solidification to segregate NCYM-binding DNA molecules from the non-binding molecules. The target beads were prepared by an amine coupling reaction using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC; Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions and washed with the selection buffer [SB; 40 mM HEPES (pH 7.5), 125 mM NaCl, 5 mM KCl, 1 mM MgCl<sub>2</sub>, and 0.01% Tween 20]. Briefly, the CA magnetic beads were washed twice with 500  $\mu$ l of 100 mM 2-Morpholinoethanesulfonic acid mono hydrate (MES) buffer at pH 6.0. After added 50  $\mu$ l of 100 mM MES buffer and 50  $\mu$ l of EDC, and the mixture was incubated for 30 minutes at room temperature. The mixture was then mixed with 200  $\mu$ g of NCYM in 100 mM MES buffer and incubated overnight at room temperature to react with the amino group of NCYM and the carboxylic acid of the CA magnetic beads. Next, they were washed twice with 500  $\mu$ l of PBST (0.1% Tween20 in phosphate buffered saline, PBS) and 500  $\mu$ l of PBST-BSA (0.1% bovine serum albumin in PBST) was added. Before using the target beads, they were washed twice with 1 ml of SB.

An initial single-stranded DNA (ssDNA) pool, 5'-GGAATGTGGTCCCTCGCAATAAATC-(N30)-GAAATGAGCCCTTTGACCCTGTAC-3', containing 30 random nucleotides between forward (Fw) and reverse (Rv) primer region was purchased from Integrated DNA Technologies (Tokyo, Japan). The selection of aptamers was performed starting from 4.5 nmol of DNAs ( $\sim 10^{15}$  molecules) in 100  $\mu$ l of SB. The pool was mixed for 15 minutes with 250  $\mu$ g of target beads at 25°C. The beads were then washed with SB, and the bound ssDNA was eluted with 7 M urea. After recovery of the eluted ssDNA using Rv primer beads, polymerase chain reaction (PCR) was performed with KOD Dash DNA polymerase (Toyobo, Osaka, Japan), a forward (Fw) primer (5'-GGAATGTGGTCCCTCGCAATAAATC-3') and a reverse (Rv) primer (5'-GTACAGGGTCAAAGGGCTCATTTTC-3') with modification by the 5'-biotin. Next, the amplified double stranded DNA (dsDNA) was bound to MyOne SA C1 magnetic beads, and the Fw chain (ssDNA) was eluted with 0.02M NaOH. The ssDNA was used for the next round.

After eight rounds of selection, the frequency of ssDNA sequences was determined by next-generation sequencing (NGS) from rounds 3 to 8 of SELEX using a MiniSeq System (Illumina, San Diego, CA, USA). Sequencing data were preprocessed by using the program of PRINSEQ++ (20) and adopted above 99.9% of the base calling accuracy (Q score of 30 and above).

### 2.2 Bio-layer interferometry

All the Bio-Layer Interferometry (BLI) measurements were performed at 25°C using an Octet<sup>®</sup> RED96e system (Sartorius AG, Goettingen, Germany). All the samples were placed in a 96 well microplate and the sample volume was 200  $\mu$ l/well. The microplate was shaken at 1,000 rpm during the measurement. As a ligand, each aptamer with 5'-biotin modification was immobilized on an Octet<sup>®</sup> SA biosensor chip (Sartorius). For kinetics analysis, different concentrations of NCYM (25–400nM) were used. The dissociation constants between the aptamer and NCYM were calculated using a

simple 1:1 biomolecular interaction model according to the manufacturer's instructions.

## 2.3 Construction of DNA frame structures and introduction of aptamers for AFM imaging

To gain insights into the mechanism of interactions between NCYM and the aptamers obtained by SELEX, NCYM-aptamer binding was studied at the single molecule level using AFM. Scaffolds prepared using the DNA origami method were used for this purpose. The DNA origami method allows the creation of structures of any shape and the introduction of functional molecules anywhere in the structure. Therefore, single-molecule observation using DNA origami structures is suitable for evaluating biomolecules and has been used to observe various molecules (21–29). In this study, we used a DNA frame structure (21). The DNA frame contains a space inside, and dsDNA can be introduced into any sequence.

DNA frames were prepared as previously described (21). A solution containing 10 nM M13mp18 ssDNA (tilibit nanosystems GmbH, Germany), 25 nM staples (2.5 eq), 20 mM Tris-HCl (pH 7.4), 10 mM MgCl<sub>2</sub>, 1mM EDTA was prepared and annealed at a rate of -1°C per minute from 85°C to 15°C.

Aptamers and DNA oligos for the DNA frame were purchased from Eurofins Genomics K.K. (Tokyo, Japan) and used without further purification. In this study, three types of aptamers, which showed high affinity among the aptamers obtained by SELEX, were employed, and the sequences of which for AFM observation are as follows:

No. 1: 5'-GGAATGTGGTCCCTCGCAATAAATCTATGTA  
CGTTATTCCTTTGACC  
AATGCTGAAATGAGCCCTTTGACCCTGTAC  
TTTTTTTCCAGCGGGACTAGCGCGTTGCTC  
CTCACT-3'

No. 2: 5'-GGAATGTGGTCCCTCGCAATAAATC  
GGGAGGGAGGGTGGGGCGGT  
GGGAGGTGAAATGAGCCCTTTGACCCTGTAC  
TTTTTTTCCAGCGGGACTAGCGCGTTGCTCCT  
CACT-3'

No. 3: 5'-GGAATGTGGTCCCTCGCAATAAATCGGGCG  
TTGTGGAGGGGGCGGTG  
GGTGGGGGAAATGAGCCCTTTGA  
CCCTGTACTTTTTTCCAGCGGG  
ACTAGCGCGTTGCTCCTCACT-3'

All aptamers had a dsDNA complement sequence (underlined) added via a TTTT sequence (bold) at the 3' end for introduction into the DNA frame.

The secondary structures of the aptamers were predicted using RNAfold [<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>]. We used the default settings: minimum free energy (MFE) and partition function, yes; avoid isolated base pairs, yes;

incorporate G-Quadruplex formation into the structure prediction algorithm, yes; dangling end options, dangling energies on both sides of a helix in any case; energy parameters, DNA parameters (Mathews model, 2004) at 25 or 38°C with a salt concentration of 1.021M.

The resultant dsDNAs were incorporated into the frame structures described above and observed in the presence of NCYM (Figure S1), which was purified according to a previous method (1).

## 2.4 Truncation and mutation of the DNA aptamer No. 1

To identify where the aptamers interact with NCYM, we prepared three analogs (short 1, short 2, and short 3) of the No. 1 aptamer based on the secondary structure prediction by RNAfold. The sequences used are as follows:

short 1: 5'-TGGTCCCTCGCAATAAATCTATGTA  
CGTTATTCCTTTGACCAATGC  
TGAAATGAGCTTTTTCAGCGGGACTAGCG  
CGTTGCTCCTCACT-3'

short 2: 5'-GGAATGTGGTCCCTCGCAATAAA  
TCTATGTACGTTATTCCTTTGAC  
CAATGCCCTTTGACCCCTGTACTTTT  
TTCCAGCGGGACTAGCGCGTTGCTCCTCACT-3'

short 3: 5'-GGAATGTGGTCCCTCGCATCCCC  
TTTGACCAATGCTGAAATGAGCCCT  
TTGACCCCTGTACTTTTTCAGCGGGAC  
TAGCGCGTTGCTCCTCACT-3'

short 1, short 2, and short 3 lack the 5' end and 3' end side (named "5'-3'-end" herein), the 3' end stem loop, and the central stem loop of the No. 1 aptamer, respectively.

In addition, mutations were introduced into aptamer No.1 without changing its secondary structure. The following sequences are mutants of aptamer No.1, and italic font indicates the introduced mutations.

No.1 mut 1 5'-GGAATGTGGTCCCTCGC  
CGCGCATCTATGTACGCGCGCCCTTTGACCAAT  
GCTGAAATGAGCCCTTTGACCCCTGT  
ACTTTTTTCCAGCGGGACTAGCGCGTTGCTCCTCACT-3'

No.1 mut 2 5'-GGAATGTGGTCCCTCGCAATAAATCTATG  
TACGTTATTCCTTTGACCGGACCAATGCTGAA  
ATGAGCCCTTTGACCCTGTACTTTTTTCCAGCGGGA  
CTAGCGCGTTGCTCCTCACT-3'

No.1 mut 3 5'-GGAATGTGGTCCCTCGCAATAAATCTATGT  
ACGTTATTCCTTTGACCAATGCTGAAAT  
GAGCCC*GCGG*ACCCTGTACTTTTTTCCAGCGGGAC  
TAGCGCGTTGCTCCTCACT-3'

No.1 mut 123 5'-GGAATGTGGTCCCTCGCCGCGC  
ATCTATGTACGGCGCGCCCCCGGACCAATGCTGAAA  
TGAGCCCGCGGACCCTGTACTTTTTTTCCAGCGGGAC  
TAGCGCGTTGCTCCTCACT-3'

## 2.5 AFM imaging

The AFM images were acquired using an AFM system (NanoWizard UltraSpeed, JPK) equipped with a silicon nitride cantilever (Olympus, BL-AC40TS). In all the measurements, NCYM (100 nM) 10eq was added to the adjusted frame structures having aptamer (10 nM) and incubated at 25°C for 2h. Samples were then double-diluted in annealing buffer (20 mM Tris-HCl (pH 7.4), 10 mM MgCl<sub>2</sub>, 1mM EDTA), adsorbed onto fresh mica plates for 10 minutes at room temperature, and washed three times with the annealing buffer. The observations were performed using the same buffer.

## 2.6 Immunoprecipitation

20 µl of Dynabeads™ protein G (Thermo Fisher Scientific) slurry was transferred to a clean tube. The tube was placed in a magnetic separation rack for 10-15 seconds, then, the buffer was carefully removed. 2 µg of anti-GSK3β (BD Transduction Laboratories) or mouse IgG (Cell Signaling Technology, Danvers, MA) was dissolved in 200 µl of PBS (0.02% Tween 20) and the solution was added to the beads. The slurry was incubated with rotation at room temperature for 10 minutes. The beads were pelleted using a magnetic separation rack and washed three times with 200 µl of PBS, then resuspended in 100 µl of PBS. 0.17µg (3.7 pmol) of purified GSK3β (Signal Chem, Richmond, Canada), 0.044g (3.7 pmol) of NCYM (1) and DNA aptamers (1eq (3.7 pmol) or 5 eq (18.5 pmol)) were dissolved in 500µl of ice-cold PBS and incubated with rotation at 4°C for 2h. A 100 µl suspension of antibody-conjugated beads was added and incubated for 2h. The beads were pelleted using a magnetic separation rack and washed five times with 1 ml of PBS. The pellet was resuspended in 20 µl of 1X sample buffer and heated at 95°C for 5min. The beads were pelleted using a magnetic separation rack and the supernatant was subjected to the Abby analysis.

## 2.7 Abby analysis

The NCYM and GSK3β protein levels were measured using a capillary electrophoretic-based immunoassay (the Abby instrument; ProteinSimple, San Jose, CA, USA), according to the manufacturer's protocol. Briefly, the samples were combined with 0.1× sample diluent buffer and 5× fluorescent master mix denaturing buffer to acquire 0.8 µg/µl loading concentration. Subsequently, the samples were denatured for 5 min at 95 °C. The primary antibodies used in this study were anti-NCYM (1) and anti-GSK3β (#9315, Cell Signaling Technology, Danvers, MA,

USA). The Abby measurements were performed using a 12–230 kDa separation module with 25-min separation at 375 V, 10-min blocking, 30-min primary antibody incubation, and 30-min secondary antibody incubation (DM-001, ProteinSimple, San Jose, CA, USA). At the end of the run, the chemiluminescent signal was displayed as a virtual blot-like image and an electropherogram based on the molecular weight using Compass (ProteinSimple, San Jose, CA, USA).

## 2.8 Sample preparation for SAXS measurements

The expression and purification procedures for NCYM were modified from those described previously (8). Recombinant NCYM with glutathione S-transferase (GST) at the N-terminus was expressed in *Escherichia coli* BL21 (DE3) cells using the pGEX-6p-1 vector. The cells were cultured in Luria broth medium containing 0.1 mg/ml ampicillin at 30°C. When OD<sub>600</sub> reached 0.6, protein expression was induced by 0.1 mM of isopropyl-β-D-thiogalactopyranoside, and culture was continued for 6 h at 30°C. The cells were harvested by centrifugation (3,890×g, 15 min, 4°C), and then stored at –30°C until purification.

The frozen cell pellets were thawed and resuspended in phosphate-buffered saline (PBS) supplemented with cComplete EDTA-free Protease Inhibitor Cocktail (11873580001, Roche, Basel, Switzerland) and lysed by repeated sonication in ice water. The cell lysate was centrifuged at 20,000×g, 4°C for 20 min, then the supernatant was loaded onto GSTrap FF column (17513102, Cytiva, Marlborough, MA, USA) equilibrated with PBS, using a peristaltic pump. After washing out the unbound materials with PBS, GST-tagged NCYM (GST-NCYM) was eluted with an elution buffer (50 mM Tris-HCl and 10 mM reduced glutathione, pH 8.0). The flow-through was reloaded onto the re-equilibrated column, and the eluate was collected once more to increase the final protein yield.

The eluted GST-NCYM solution was dialyzed against a buffer containing 50 mM Tris-HCl, 100 mM NaCl, and 1 mM EDTA at pH 8.0. After adding dithiothreitol to the protein solution at a final concentration of 1 mM, the GST-tag was cleaved with 50 U/L-culture of PreScission Protease (27084301, Cytiva) for over 18 h at 4°C with gentle stirring. To exhaustively degrade the remaining nucleic acids, 6,000 U/L-culture of benzonase (71205-3CN/70746-3CN, Merck, Darmstadt, Germany) was added to the cleaved sample along with MgCl<sub>2</sub> at a final concentration of 2 mM (Mg<sup>2+</sup> is required for the activation of benzonase), after which the sample was dialyzed against IEX buffer (20 mM MOPS (pH 7.0) and 1 mM DTT) with 2 mM MgCl<sub>2</sub>.

The dialyzed sample was loaded onto a HiTrap SP HP column (17115201, Cytiva) equilibrated with IEX buffer, and eluted using a linear gradient of NaCl (0–650 mM). The fractions containing high-purity NCYM, as confirmed by SDS-PAGE, were collected and used for SAXS measurements. The GST-tag did not bind to the column and was detected in the flow-through. The final NCYM yield was 1.25 mg/L-culture, which was estimated with the molecular absorption coefficient of 280 nm  $\epsilon_{280}^{0.1\%} = 0.558$ .



The purified NCYM was dialyzed against a buffer containing 10 mM Tris-HCl (pH8.0), 50 mM NaCl, and 5 mM DTT. Powdered DNA aptamers (No. 1) were dissolved directly in the same buffer. Samples of the NCYM-DNA complex were prepared by mixing these two kinds of solutions at the appropriate molar ratios (see below). These solutions were used for the following SAXS measurements.

## 2.9 SAXS experiment

SAXS measurements were carried out at BL40B2 in SPring-8 (Hyogo, Japan) on solution samples of NCYM (1.4 and 2.8 mg/ml), the DNA aptamer (1.6 and 5.0 mg/ml), and the NCYM-DNA complex. For the measurements of the complex, two kinds of samples, where the molar ratio of NCYM and DNA was 1:1.2 or 1:1.5 (Table 1), were used to extract the scattering curve of the complex by changing the relative contribution of the unbound DNA aptamers. The wavelength ( $\lambda$ ) of the incident X-ray was 1.0 Å and the temperature was 293 K with the sample-to-detector distance of 2.2 m. A pixel detector (PILATUS3S 2 M, Dectris) was used to record the scattering patterns.

Data reduction was conducted using the software SAnGler (30): The recorded two-dimensional SAXS patterns were circularly averaged to obtain one-dimensional scattering curves, corrected by the incident flux measured with an ion chamber placed upstream of the samples. The net scattering curves of the scattering particles were obtained by subtracting the scattering curves of the buffer from those of the samples with an appropriate scaling factor based on the scattering particle concentration and the partial specific volumes of proteins (0.73 cm<sup>3</sup>/g) or of DNA (0.53 cm<sup>3</sup>/g) (31). Finally, the scattering curves were normalized to the absolute scale using H<sub>2</sub>O as the standard (32) to estimate the molecular weight of the scattering particles.

Guinier analysis was employed to evaluate the radius of gyration ( $R_g$ ) of the scattering particle from its scattering curve. A scattering curve  $I(Q)$ , where  $Q (=4\pi\sin\theta/\lambda$ , where  $2\theta$  is the scattering angle) denotes the momentum transfer, is represented as follows in a good approximation in the so-called Guinier region ( $Q \cdot R_g < 1.3$ ) (33):

$$I(Q) = I(0)\exp\left(-\frac{1}{3}R_g^2Q^2\right) \quad (1),$$

where  $I(0)$  [cm<sup>-1</sup>] denotes the forward scattering intensity, from which the molecular weight (MW) of the scattering particle is

estimated from the following equation in kDa (32):

$$MW = 1500 \times I(0) \frac{1}{c} \quad (2),$$

where  $c$  is the weight concentration [g/l].

For rod-like particles such as the DNA aptamer and the NCYM-DNA complex, cross-sectional Guinier analysis was applied to evaluate their cross-sectional radii of gyration ( $R_c$ ). In this case,  $I(Q)$  is approximated as:

$$Q \cdot I(Q) = I_c(0)\exp\left(-\frac{1}{2}R_c^2Q^2\right) \quad (3),$$

where  $I_c(0)$  denotes the forward scattering intensity of the cross-section of the scattering particle. Application of Eq. 2 with  $I_c(0)$  instead of  $I(0)$  yields the scattering mass per unit length.

The scattering curves taken at the lower particle concentrations were merged with those at the higher concentrations at  $Q = 0.1 \text{ \AA}^{-1}$  and the resultant curves were used for structural modeling. IGOR Pro software (WaveMetrics, Lake Oswego, OR, USA) was used for (cross-sectional) Guinier analyses and for processing the scattering curves.

## 2.10 Structural modeling using the *ab initio* method

Human NCYM comprises 109 residues. Because NCYM was expressed with a GST-tag at its N-terminus in this study, eight residues were added to the 109 residues even after cutting the tag. For structural modeling, the program GASBOR (34) was employed, where each residue is represented by a sphere with a constant electron density. The obtained model thus consists of 117 spheres. In this study, the GASBOR runs were repeated 10 times (i.e., 10 best-fit models were obtained). For the modeling of the DNA aptamer, DAMMIF (35) was used, where a molecule is represented by an ensemble of spheres called dummy atoms. The input files required for DAMMIF were generated using AUTORG and DATGNOM (36). The maximum  $Q$  value ( $Q_{\max}$ ) used for structural modeling was automatically determined using DATGNOM ( $Q_{\max}R_g < 7-8$ ). The DAMMIF runs were repeated 10 times and the resultant 10 models were averaged, followed by filtration using DAMAVER (37). Structural modeling of the NCYM-DNA complex was carried out using the program MONSA (38), where each of the two phases is represented by a dummy atom model while each phase is assigned a designated electron density value. As an input file of MONSA, the following information is required: The values of the electron density of NCYM and the DNA aptamer were set to be 0.09 e/Å<sup>3</sup> and 0.21 e/Å<sup>3</sup>, respectively, which are typical of these types of molecules (39). The volume fractions of NCYM and the DNA aptamer were obtained from the volumes obtained using GASBOR and DAMMIF, respectively. Using these parameters, the MONSA runs were repeated 10 times. In all of the programs above, the best-fit model, the scattering curve of which reproduces the experimental curve well, is determined by simulated annealing.

TABLE 1 Information on the measured samples of the NCYM-DNA complex.

Molar ratio (NCYM : DNA) in the sample	NCYM concentration in the sample [mg/ml]	DNA concentration in the sample [mg/ml]
1:1.2	1.3	3.2
	0.6	1.5
1:1.5	0.6	1.8

## 2.11 Scattering curve of the NCYM-DNA complex

The scattering curves of the NCYM-DNA complex recorded at 1:1.2 and 1:1.5 molar ratios were found to be superimposable despite the existence of different amounts of unbound DNA aptamers (Figure S2). However, as described in the Results section, NCYM molecules and the DNA aptamers exist as monomers and dimers, respectively, in solution so that the molar ratios of NCYM and the dimeric DNA aptamer in the above samples are 1:0.6 and 1:0.75, respectively, which result in the molar ratios of the complex and the excess amount of NCYM of 0.6:0.4 and 0.75:0.25 assuming that one NCYM monomer binds to a DNA dimer. Because the scattering intensity is proportional to the square of the product of the scattering contrast of a particle and its volume, the scattering contributions of excess NCYM molecules were calculated to be 0.8% and 0.4% for the 0.6:0.4 and 0.75:0.25 samples, respectively. In this estimation, the scattering contrasts of NCYM and DNA were set to be  $0.09 \text{ e}/\text{\AA}^3$  and  $0.21 \text{ e}/\text{\AA}^3$  as described above, and the volumes of a NCYM monomer and a DNA dimer were assumed to be  $17000 \text{ \AA}^3$  and  $68000 \text{ \AA}^3$  as obtained by GASBOR and DAMMIF. Thus, it appears that the scattering contribution of free NCYM monomers is negligible, which explains why the scattering curves obtained at NCYM : DNA = 1:1.2 and 1:1.5 are similar to each other within errors. This observation also excluded the possibility that the complex consists of one NCYM monomer and one monomeric DNA aptamer. Based on the above inspection, the curves obtained from the above samples reflect those of the NCYM-DNA complex. The scattering curves taken at NCYM : DNA = 1:1.2 and 1:1.5 were, therefore, averaged to improve the signal-to-noise ratio and the resultant curve was employed as the scattering curve of the NCYM-DNA complex.

## 3 Results

### 3.1 Aptamer selection by SELEX

To identify DNA aptamers that specifically bind to NCYM, we performed SELEX and employed NGS technology to monitor the progress of the enrichment sequences that bind to the target in the selection pool. We performed the NGS analysis using an Illumina MiniSeq for the NCYM aptamer selection pools from rounds 3 to 8 to identify the aptamer candidates. The ratio of unique DNA sequences in the selection pool per round showed that DNA sequences were enriched in round 6 (Figure 1A). No. 1 aptamer candidate, which was the most abundant population in round 8, was quickly enriched from round 6 compared to the other DNA sequences. Guanine contents in the selection pools were increased slightly in round 8 (Table S1). BLI measurements and kinetic analyses showed that top three aptamers strongly bind to NCYM protein at  $K_D$  value in the range of 53.9 to 299 nM (Table S2); however, the aptamer No. 2 and 3 were predicted to form different secondary structure at 38°C, showing low structural stability at relatively high temperature compared to aptamer No.1 (Figure 1B and Figure S3).

## 3.2 AFM observation on the NCYM-DNA aptamer complexes

### 3.2.1 DNA aptamers obtained by SELEX (No. 1, 2, and 3)

To visualize the interaction between NCYM and DNA at the molecular level, we used AFM observation of the aptamer-conjugated DNA frames with or without NCYM (Figure 2). Because the aptamer has a large single-stranded region and is not completely fixed, only dsDNA can be seen. After adding NCYM, white dots were observed on the dsDNA (Figure 2B). The numbers of DNA frames with and without these white dots were counted to assess the affinity of the aptamer for NCYM (Figure 2B). Binding of NCYM to the No. 1 aptamer was observed in 14.5% of the DNA frames (84/580 frames). Binding was also observed for the No. 2 and No. 3 aptamers at 12.0% (80/664 frames) and 14.6% (41/280 frames), respectively. The slightly lower affinity of No. 2 compared to No.1 and No.3 is consistent with the  $K_D$  value evaluated by BLI measurements.

### 3.2.2 Truncated and mutated DNA aptamers

Because of the relatively higher stability of the secondary structures of the No.1 aptamer (Figure S2), we focused on aptamer No.1 and further analyzed the regions required for NCYM binding. As shown in Figure 3A, the AFM results revealed a decrease in the number of bonds in short 1 (9.4%, 70/742 frames) and short 2 (5.3%, 75/1425 frames), whereas there was no change in the number of bonds in short 3 (14.2%, 173/1222 frames). Because aptamers No. 2 and No. 3 exhibited GC-rich sequences (Figure 1B), we introduced mutations in the No. 1 aptamer with increasing GC content without affecting the secondary structure, and we found that the mutations at TTT in the central loop showed a significant decrease in the affinity of aptamer No.1 to NCYM (Figure 3B). These results suggest that the 5'-3'-end, the 3' end stem loop, and TTT in the central loop of the No.1 aptamer are required for binding to NCYM.

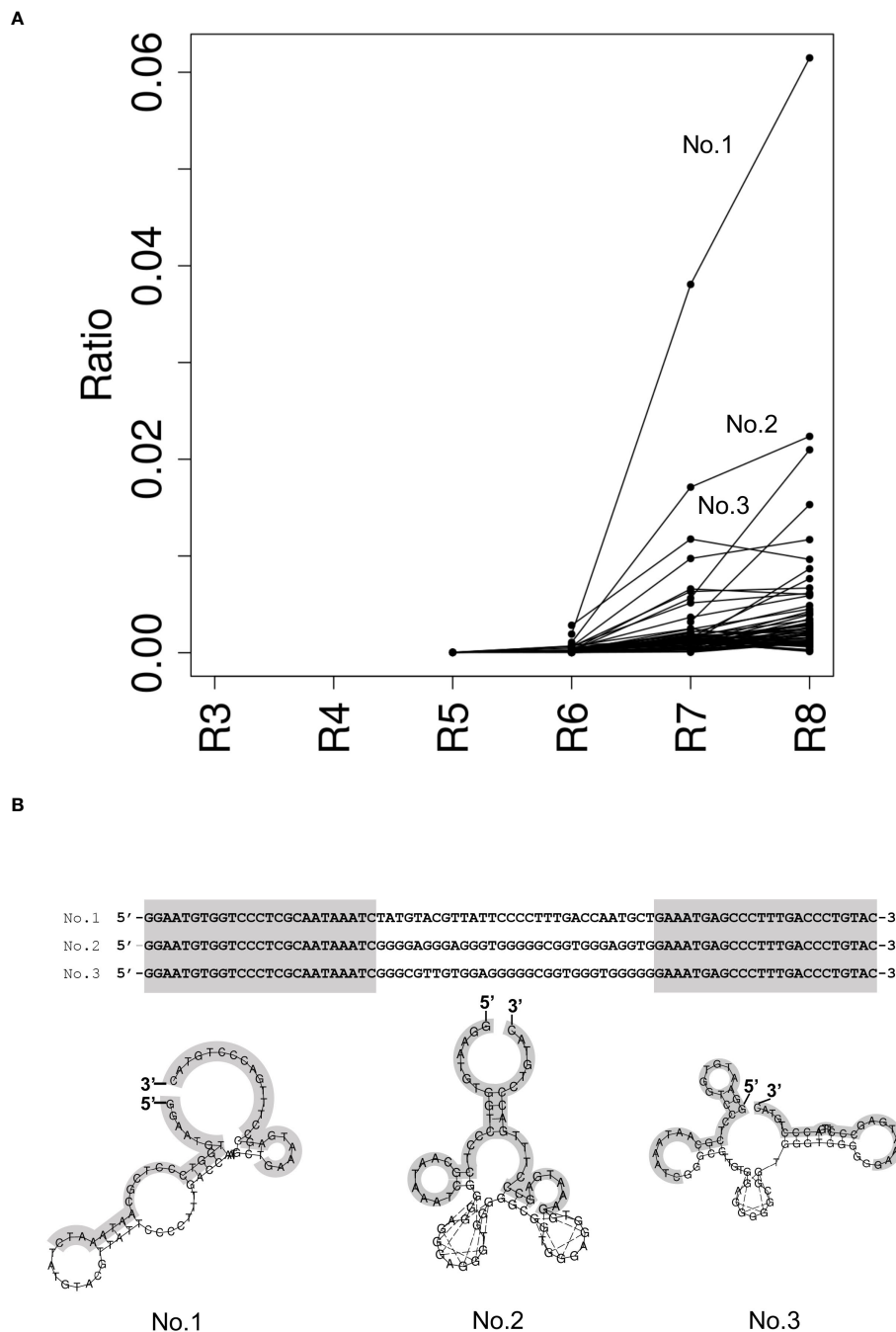
## 3.3 The effect of DNA aptamer No. 1 on the interaction between NCYM and GSK3 $\beta$

To clarify the effect of DNA aptamer No.1 on NCYM function, we examined NCYM binding to GSK3 $\beta$  with or without the aptamer. As reported (1), purified NCYM was co-immunoprecipitated with GSK3 $\beta$  (Figure 4A). Addition of aptamer No.1 enhanced the interaction between NCYM and GSK3 $\beta$  in dose-dependent manner (Figure 4B). This result led us to analyze the NCYM-aptamer No.1 complex because the aptamer may help NCYM to adopt a conformation that facilitates binding to GSK3 $\beta$ .

## 3.4 SAXS results

### 3.4.1 Guinier analysis and Kratky plot

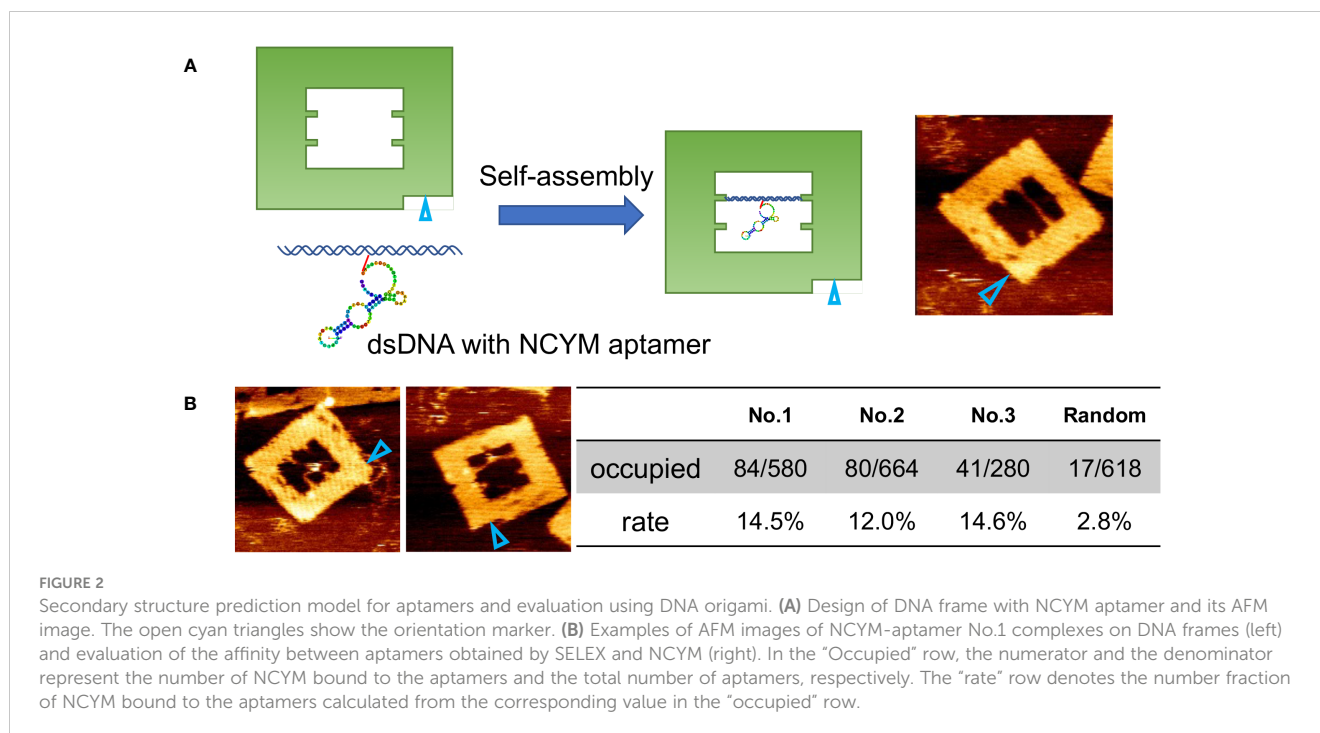
To elucidate their structural properties, we employed SAXS analysis to reveal the approximate structures of the complexes in



**FIGURE 1** Identification of NCYM-bound aptamers. **(A)** Ratio of the enriched ssDNA sequence in the selection pool per round. Each dot represents the ratio of the enriched ssDNA sequences more than 0.001. The lines indicate the same ssDNA sequence between rounds. Top 3 enrichment sequences at round 8 were evaluated for the further structural analysis and named No. 1, No. 2, and No. 3. **(B)** Secondary structure prediction of NCYM-bound aptamer by RNAfold at 25°C. Primer sequences are colored in gray.

solution. Figure 5A shows the results of the Guinier analysis. The radii of gyration ( $R_g$ ) were estimated to be  $25.2 \pm 0.7 \text{ \AA}$ ,  $46.0 \pm 0.3 \text{ \AA}$ , and  $44.8 \pm 0.2 \text{ \AA}$ , for NCYM, the DNA aptamer No. 1, and the NCYM-DNA complex, respectively. The molecular weights of NCYM and the DNA aptamer were estimated to be 7.2, and 54.5 kDa, respectively, indicating that NCYM molecules exist as a monomer in solution whereas DNA aptamers are dimers because the molecular weights of NCYM and the DNA aptamer were 12

kDa, 24.2 kDa, respectively. Although the molecular weight of the complex was estimated to be 46.8 kDa, because the true concentration of the complexes in the sample is not known due to the existence of a small amount of excess NCYM molecules, this value should be interpreted with caution. Scattering curves other than NCYM were found to follow the cross-sectional Guinier approximation, suggesting that the DNA aptamer and the complex adopt a rod-like shape. Figure 5B shows the results of



the cross-sectional Guinier analysis, from which the cross-sectional radii of gyration ( $R_c$ ) were estimated to be  $12.2 \pm 0.4 \text{ \AA}$  and  $12.5 \pm 0.2 \text{ \AA}$  for the DNA aptamer and the complex, respectively. This suggests that the overall size of the cross-section is similar between the DNA aptamer and the complex. Both the Guinier analysis and the cross-sectional Guinier analysis can be applied to the current samples because of their relatively short entire length, whereas in many cases, fibrillar proteins are quite long (even on the order of  $\mu\text{m}$ ) and only the cross-sectional Guinier analysis can be applied (40, 41).

Several plots were used to investigate the structural properties of NCYM, as shown in Figure 5C. The Kratky plot ( $Q^2 \cdot I(Q)$  vs  $Q$ ) (42) shows that the  $Q^2 \cdot I(Q)$  value reaches a peak at around  $Q = 0.1 \text{ [\AA}^{-1}\text{]}$  and decreases slightly, followed by the increase, suggesting that NCYM is not a completely unfolded protein, but a partially folded protein. This interpretation is further supported by the  $Q^3$  plot ( $Q^3 \cdot I(Q)$  vs  $Q^3$ ), where the  $Q^3 \cdot I(Q)$  value reaches a plateau, which is a hallmark of partially folded proteins (43). In contrast, no plateau was observed in the Porod-Debye plot (43), suggesting that NCYM is not a well-folded protein. These results show that NCYM adopts a highly flexible conformation, though it is not a completely unfolded protein, but a partially folded protein. This is in agreement with a previous study showing that NCYM molecules contain defined secondary structures in solution (8).

### 3.4.2 Solution structure of NCYM, the DNA aptamer, and the NCYM-DNA complex

The results of the structural modeling of NCYM, the DNA aptamer No. 1, and the NCYM-DNA complex are summarized in Figure 6. Figure 6A compares the experimental SAXS curves with those calculated using the best-fit model obtained for each of the three samples. The  $\chi^2$  values, which are averaged over 10 models

obtained, were 1.17, 1.37, and 1.85, for NCYM, the DNA aptamer, and the complex, respectively. As shown in Figure 6A, the obtained models provide excellent fits to the measured SAXS curves.

As shown in Figure 6B, NCYM was found to have a slightly extended structure in which the bulky and flexible parts are mixed. It is not a completely unfolded protein such as a random coil, as expected from the panels in Figure 5C and our previous study on the secondary structure of NCYM (8). Although the assignment of the secondary structure to the three-dimensional structures of the current NCYM models is not possible at this stage, future structural modeling using SAXS curves with higher  $Q$  values would be useful for this purpose (44). The volume of the NCYM models was  $\sim 17000 \text{ \AA}^3$ . Regarding the DNA aptamer, a structure with a volume of  $\sim 68000 \text{ \AA}^3$ , in which four to five bulky nodes were connected, was found to reproduce the measured SAXS curve, as shown in Figure 6C. Because the aptamers exist as dimers, as evidenced by the molecular weight estimation as described above, a monomeric DNA aptamer corresponds to two or three nodes in this model. The structural features of the DNA model with some nodes here are consistent with those predicted using RNAfold (Figure 1B). Two representative models of the NCYM-DNA complex are shown in Figure 6D. In both cases, the volumes of the NCYM molecule and of the DNA aptamer in the complex were  $\sim 19000 \text{ \AA}^3$  and  $\sim 63000 \text{ \AA}^3$ , which are roughly the same as those obtained when these molecules are in isolation. The volume of the complex ( $\sim 82000 \text{ \AA}^3$ ) was essentially the same as the sum of the volumes of NCYM and DNA in isolation ( $\sim 85000 \text{ \AA}^3$ ) within 4% accuracy. It thus follows that the NCYM-DNA complex consists of one NCYM molecule and two DNA aptamers (a dimer). NCYM tends to bind either to or close to a tip of a dimeric DNA aptamer. The slight differences ( $\sim 10\%$ ) in the volumes of each component between in isolation and in the complex may imply that some intramolecular structural



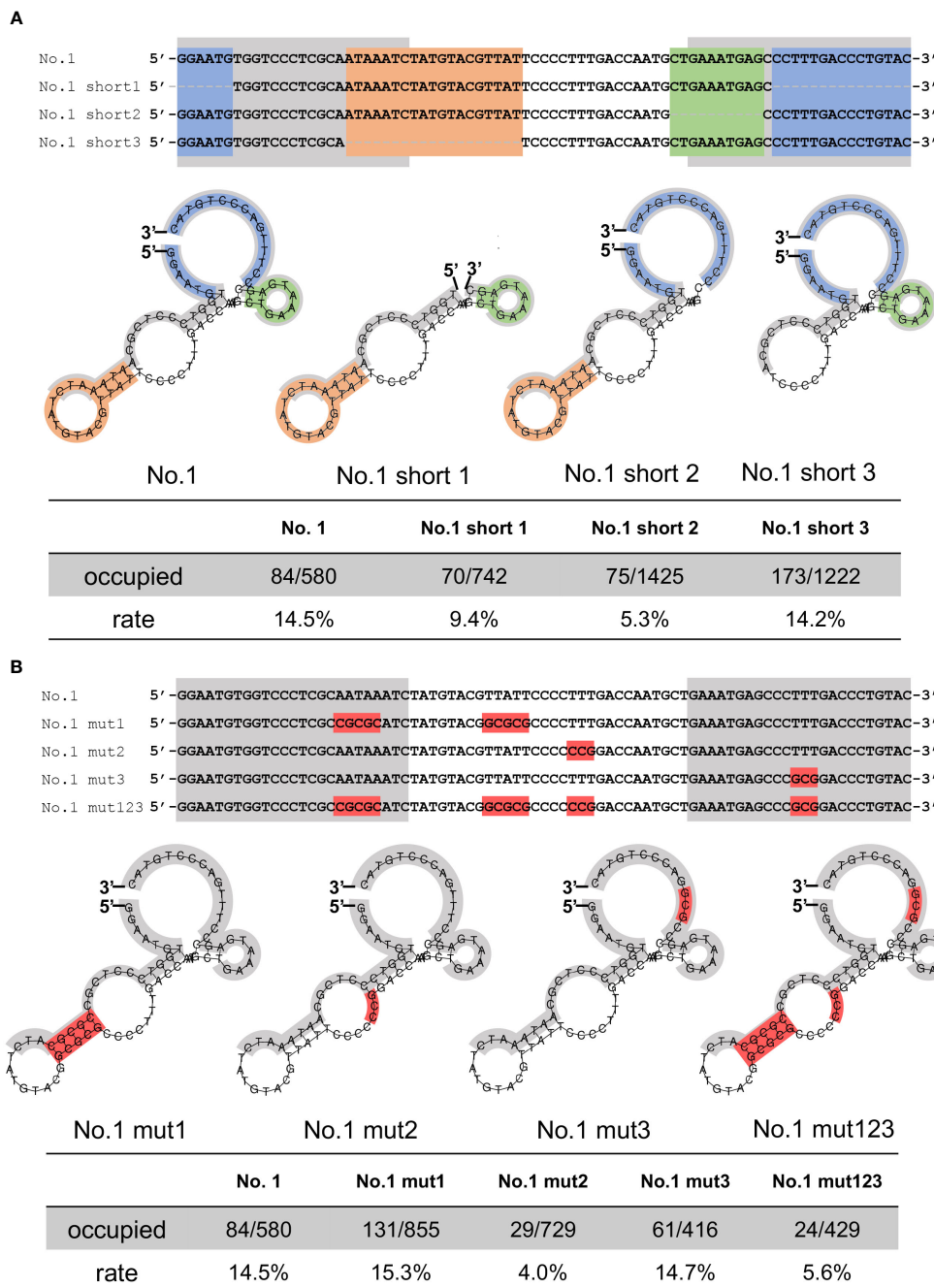


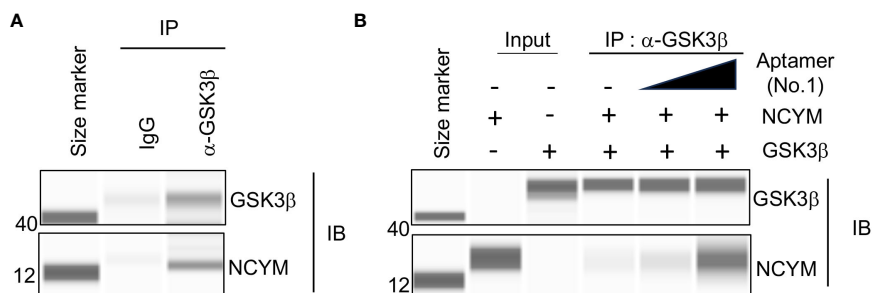
FIGURE 3

Evaluation of the affinity between truncated or mutated aptamers and NCYM. (A) Truncated No.1 aptamers for identification of NCYM binding site. The 5'-3'-end, the central stem loop, the 3' end stem loop, and the primer region are colored in blue, orange, green, and gray, respectively (upper and middle). Evaluation of the affinity between the truncated aptamers and NCYM (bottom) (B) Mutated No.1 aptamers for identification of the NCYM binding site. Mutated sequences are colored in red (upper and middle). Evaluation of the affinity between the mutated aptamers and NCYM (bottom).

changes occur upon binding. A more detailed structure of the complex can be obtained by small-angle neutron scattering (SANS) combined with contrast matching or variation (45), which will be performed in the future. All the 10 models obtained for NCYM, the DNA aptamer, and the NCYM-DNA complex are shown in Figures S4–S6, respectively.

Attempts were made to identify other possible conformations of the complex by changing the volume ratio of each phase in the

complex and/or assuming a symmetry in the structure, which are provided as an input file to MONSA. Whereas several models which fit the SAXS curve of the complex quite well in terms of the  $\chi^2$  values were obtained, in all cases, deviation of the volumes of NCYM and the DNA aptamer in the complex with regards to those in the unbound state was much larger (15–20%) than the models presented in Figure 6. Considering an independent line of evidence that the volume change of proteins between the folded and the

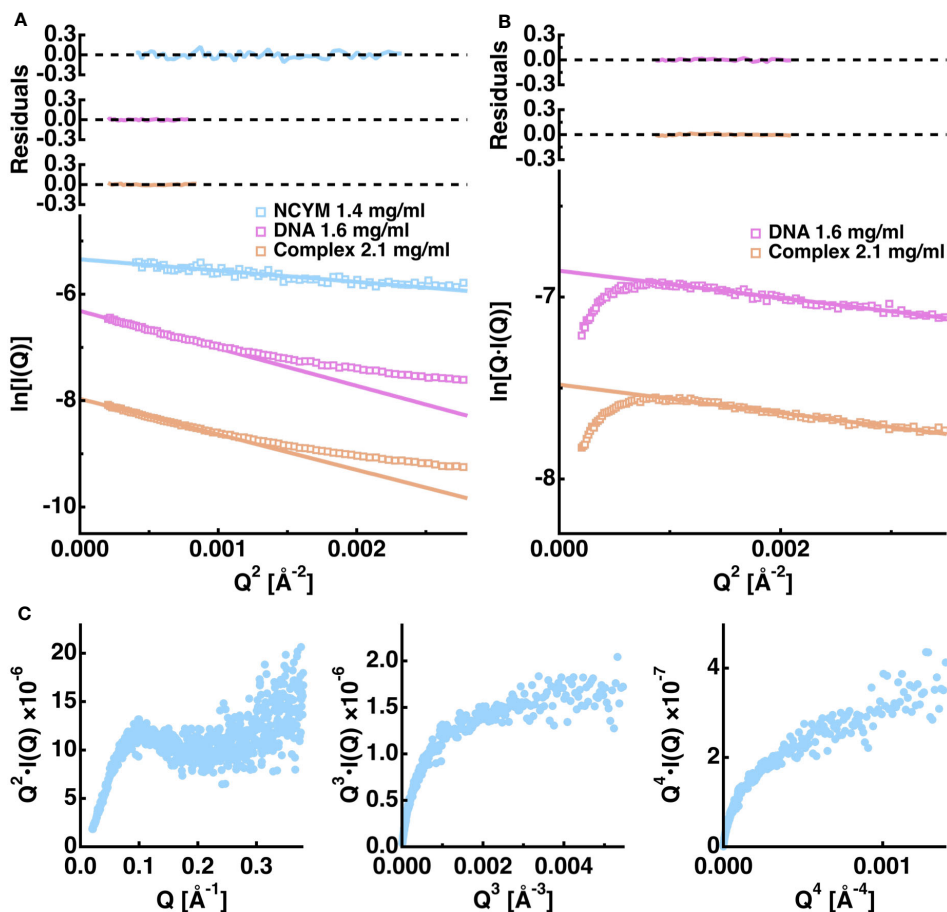


**FIGURE 4** Aptamer No.1 increased interaction between NCYM and GSK3 $\beta$ . **(A)** Co-immunoprecipitation of NCYM with GSK3 $\beta$  detected by Abby analysis. **(B)** Aptamer No.1 increased the amount of NCYM co-immunoprecipitated with GSK3 $\beta$  in a dose-dependent manner (0, 1, and 5 eq.).

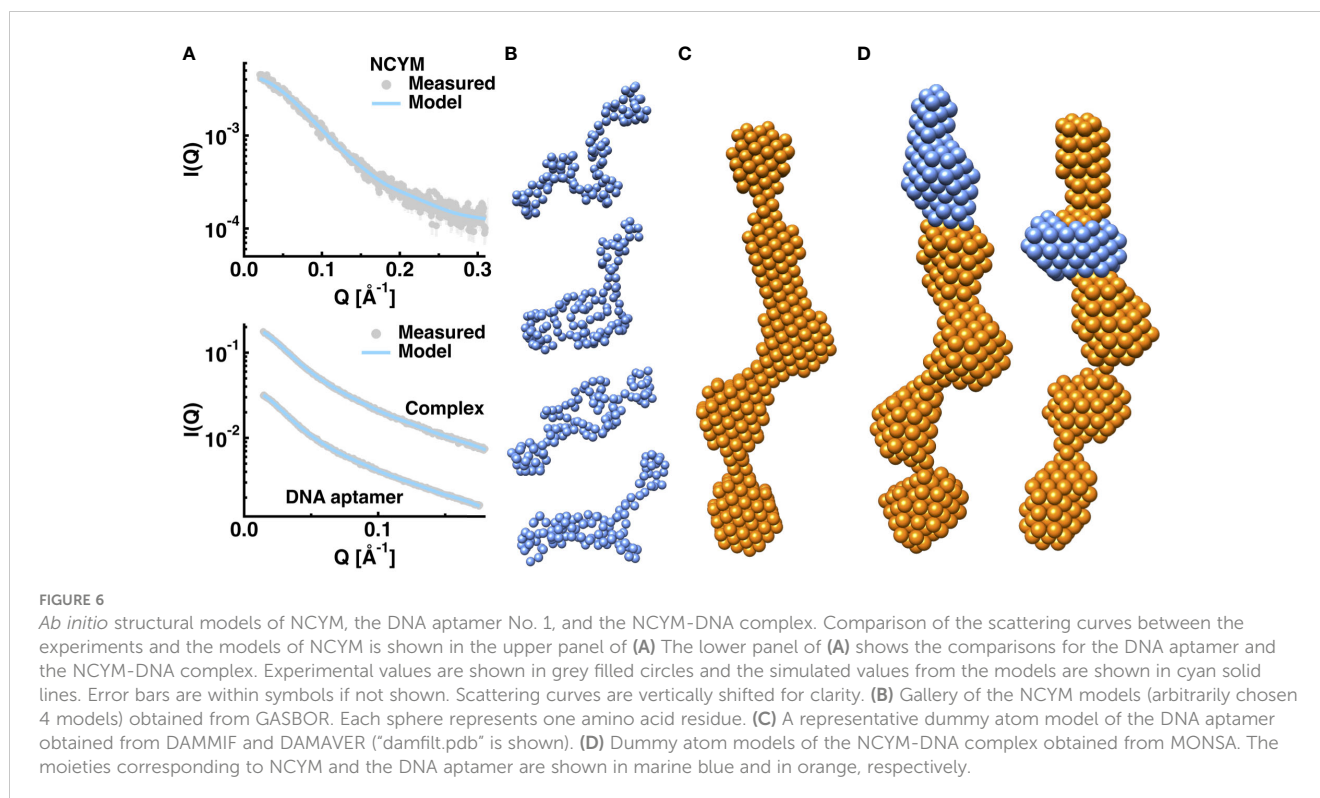
unfolded states is less than 0.5% (46), the volume change of each component in the complex should be the smallest. Since in the models in Figure 6, the volume of each component in the complex is close to that in the unbound state, and truncation of one edge of a DNA aptamer breaks down the interaction with NCYM as observed by AFM, the models presented here appear to be reasonable.

### 4 Discussion

In this study, we established a new observation system using aptamers for the single-molecule observation of proteins using DNA origami. To date, protein studies using DNA origami have mainly involved single-molecule observations of proteins that bind



**FIGURE 5** Summary of the analysis of the SAXS scattering curves of the NCYM-DNA system. **(A)** Guinier analysis. The logarithm of the scattering intensity is shown as a function of  $Q^2$  for NCYM (cyan), the DNA aptamer No. 1 (magenta), and the NCYM-DNA complex (orange). Upper panels denote the corresponding residuals between the measured and the fitted values. **(B)** Cross-sectional Guinier analysis. Instead of  $\ln[I(Q)]$  of the Guinier analysis,  $\ln[Q \cdot I(Q)]$  is plotted as a function of  $Q^2$ , from which the cross-sectional radius of gyration is evaluated. **(C)** The left, middle, and right panels show the Kratky plot,  $Q^3 \cdot I(Q)$  vs  $Q^3$  plot, and the Porod-Debye plot, respectively, of NCYM.



directly to DNA or using systems based on ligand or avidin-biotin binding (21–29, 47–52). However, DNA origami research using aptamers has mainly focused on functionalizing the DNA origami using already established aptamers (53–56). Therefore, a system that uses DNA origami and aptamers to elucidate the structure of proteins, as in this study, is a new approach that has never been used. In this study, we have also succeeded in roughly identifying the binding site of NCYM to the aptamer obtained using the SELEX method. The method using aptamers can easily fix proteins onto the DNA origami. Therefore, it is expected to be applied to the observation of the interaction between a fixed protein and its target, and to the functional evaluation of proteins for which single molecule observation has not been performed.

In addition to the AFM observation, further structural characterization of NCYM and the DNA aptamer No. 1 that facilitates NCYM binding to GSK3 $\beta$  was conducted using SAXS. Regarding the structure of the DNA aptamer, it was found that it exists as a dimer. This finding is supported by both the molecular weight estimation from the forward scattering intensity and the fitting of the corresponding SAXS curve. Although it is not possible to unambiguously determine the manner in which the two DNA aptamers form dimers, this can be inferred from the current findings. There are three types of arrangements for the two DNA aptamers to form a structure similar to that shown in Figure 6C (Figures 7A–C).

As shown in Figure 7, there are three types of arrangements for two DNA aptamers to form a similar form to that shown in Figure 6C. In the case where the interface between the two aptamers is formed by the central stem loop of one aptamer and the 5'-3'-end of the other aptamer (Figure 7A), NCYM is able to

bind to the 5'-3'-end of the upper aptamer. Whereas there is a possibility that other aptamers bind to either edge of the dimer in the same manner, dimers might be more stable than higher-order aggregates as observed for a protein (57) probably due to entropy-enthalpy compensation. In case where two 5'-3'-ends form the interface of the dimer (Figure 7B), NCYM would not be able to bind to the aptamer since these regions are required for NCYM binding as suggested by AFM. If the interface of the dimer is formed by two central stem loops (Figure 7C), at least two NCYM molecules would be able to bind to both edges of the dimer, which is inconsistent with the discussion on the SAXS curve of the complex and our modeling results. The second NCYM binding may be unfavorable in terms of entropy. There are thus two possible models on the mode of dimeric formation of the DNA aptamer (Figures 7A, C). In addition, the observation that NCYM binds to the dimeric DNA aptamer (Figure 6) implies that the affinity between NCYM and an aptamer is lower than that between monomeric aptamers and thus DNA dimers do not dissociate into two monomeric aptamers in the current solution condition.

A concern in the current SAXS analysis is that the deviation of the estimated molecular weight (7.2 kDa) of NCYM from its theoretical value (11.7 kDa) is relatively large. This raises the possibility that the DNA aptamers might exist as trimers or tetramers. The large deviation observed for NCYM is likely to be caused by the small size of NCYM, thus resulting in relatively large experimental errors. However, because the scattering intensity is proportional to the square of the molecular volume, the SAXS curve of the DNA aptamer, which has a larger molecular weight than NCYM, has much lower experimental errors than NCYM (Figure 6A). This results in a more reliable molecular weight

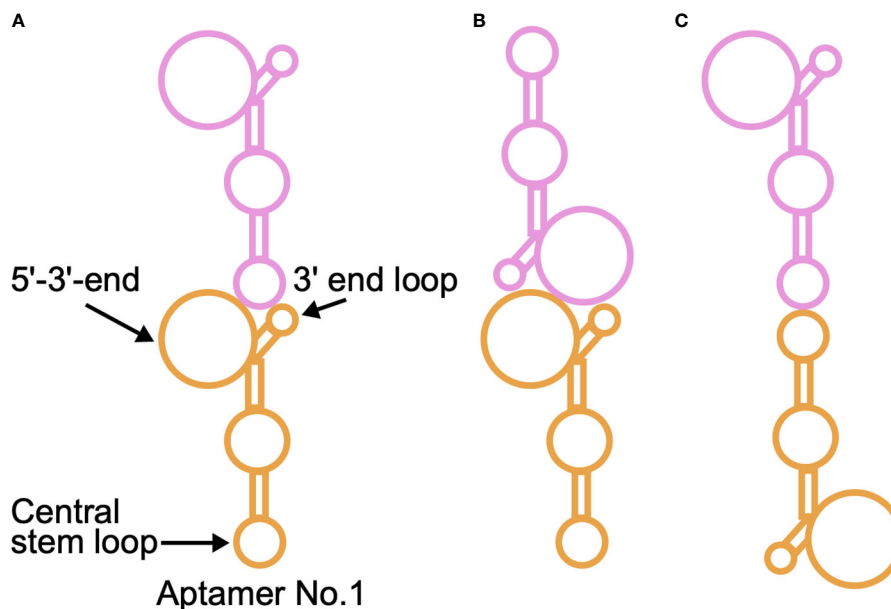


FIGURE 7

Schematic illustration of possible ways by which the DNA aptamer No. 1 forms a dimer. An aptamer is shown in either orange or magenta. The corresponding predicted structure of the aptamer is shown in Figure 1B (left). There are three types of arrangements for two DNA aptamers to take a form similar to that shown in Figure 6C: (A) The interface between the two aptamers is formed by the central stem loop of one aptamer and the 5'-3'-end of the other aptamer. (B) Two 5'-3'-ends form the interface of the dimer. (C) The interface of the dimer is formed by two central stem loops. For a more detailed discussion, please see the main text.

estimation for the DNA aptamer than for NCYM. Furthermore, the structural modeling of an isolated DNA aptamer presented in this work does not require any information on the molecular property including its molecular weight because it is an “ab initio” approach. From the SAXS-derived models with five bumps (Figure 6) and the secondary-structure predicted by RNAfold, which has 2–3 bumps in one aptamer (No. 1 in Figure 1B), it is reasonable to conclude that the DNA aptamer forms a dimer in solution.

In the dummy atom models of the complex, the moiety corresponding to NCYM adopted a compact and well-defined shape whereas NCYM showed flexibility and adopted a slightly extended conformation in the unbound state. This implies that highly flexible NCYM molecules fold upon binding to DNA through the well-known “fly-casting mechanism” (58), in which an unfolded region(s) of a protein binds weakly to the binding site at a relatively large distance, followed by folding as the protein approaches the binding site. As aptamer No.1 facilitated the interaction between NCYM and GSK3 $\beta$ , the compact and well-defined shape of NCYM found in the complex with aptamer No. 1 appears to be the functional structure of NCYM. Folding upon binding has been demonstrated for another *de novo* protein AFGP8 (16). A similar mechanism may also apply to interactions between NCYM and GSK-3 $\beta$ , underlying the mechanism of stabilization of these molecules.

One of the most important advantages of AFM observations with a DNA frame is that the DNA frame is guaranteed to be a monomeric aptamer, which is a feature not observed in other

binding assays. This is because dsDNA is bound to a DNA aptamer (Figure 2A) and thus two dsDNAs should be observed by AFM if the DNA aptamers form dimers, which is not the case. Therefore, it is most likely that monomeric aptamers are capable of binding NCYM and dimer formation of DNA aptamers is not a prerequisite for binding of NCYM. Since the SAXS data alone did not show that NCYM could bind to the monomeric form of the aptamer, AFM, in combination with SAXS, provides important insights into the structure of the aptamer-NCYM complex.

Identification of DNA aptamers that can specifically bind to NCYM and facilitate its binding to GSK3 $\beta$  is useful for elucidation of the structure of NCYM in its active form. Although the aptamer-NCYM complex described in this study has not yet been tested by X-ray crystallography, investigation of the structural properties of the complexes may pave the way for the structural characterization of NCYM molecules at the atomic level. Therefore, the present study suggests that the combination of SELEX, AFM, and SAXS is useful for understanding the structural properties of NCYM and the current findings will serve as a foundation for the future design of novel anti-cancer drugs targeting NCYM as well as for elucidating the stabilization mechanism of other cancer-related proteins by NCYM.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.



## Author contributions

YS and TM designed and supervised the study. SY, FK, KN, MH, KH, TT, YH, YS and TM performed the experiments, acquired, and analyzed the data. SY, FK, KN, TT, YS and TM wrote the manuscript. KN, YH, TT, YS and TM acquired funds. All authors contributed to the article, read, and approved the submitted version.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the AMED eASIA JRP (22jm0210092h0002), and by the JSPS Bilateral Program Number JPJSBP120233212 (SAKURA project).

## Acknowledgments

The authors thank SPring-8 for the allocation of the beamtime. The SAXS experiments were carried out under the approval of the SPring-8 Program Review Committee (2022B1242).

## References

- Suenaga Y, Islam SMR, Alagu J, Kaneko Y, Kato M, Tanaka Y, et al. NCYM, a cis-antisense gene of MYCN, encodes a *de novo* evolved protein that inhibits GSK3 $\beta$ . Resulting in the stabilization of MYCN in human neuroblastomas. *PLoS Genet* (2014) 10:e1003996. doi: 10.1371/journal.pgen.1003996
- Suenaga Y, Nakatani K, Nakagawara A. *De novo* evolved gene product NCYM in the pathogenesis and clinical outcome of human neuroblastomas and other cancers. *Jpn J Clin Oncol* (2020) 50:839–46. doi: 10.1093/jjco/hyaa097
- Shoji W, Suenaga Y, Kaneko Y, Islam SMR, Alagu J, Yokoi S, et al. NCYM promotes calpain-mediated Myc-nick production in human MYCN-amplified neuroblastoma cells. *Biochem Biophys Res Commun* (2015) 461:501–6. doi: 10.1016/j.bbrc.2015.04.050
- Suenaga Y, Yamamoto M, Sakuma T, Sasada M, Fukai F, Ohira M, et al. TAp63 represses transcription of MYCN/NCYM gene and its high levels of expression are associated with favorable outcome in neuroblastoma. *Biochem Biophys Res Commun* (2019) 518:311–8. doi: 10.1016/j.bbrc.2019.08.052
- Kaneko Y, Suenaga Y, Islam SMR, Matsumoto D, Nakamura Y, Ohira M, et al. Functional interplay between MYCN, NCYM, and OCT4 promotes aggressiveness of human neuroblastomas. *Cancer Sci* (2015) 106:840–7. doi: 10.1111/cas.12677
- Islam SMR, Suenaga Y, Takatori A, Ueda Y, Kaneko Y, Kawana H, et al. Sendai virus-mediated expression of reprogramming factors promotes plasticity of human neuroblastoma cells. *Cancer Sci* (2015) 106:1351–61. doi: 10.1111/cas.12746
- Zhu X, Li Y, Zhao S, Zhao S. LINC15A activates Wnt/ $\beta$ -catenin signaling by interacting with NCYM to promote bladder cancer progression. *Biochem Biophys Res Commun* (2018) 502:299–306. doi: 10.1016/j.bbrc.2018.05.076
- Matsuo T, Nakatani K, Setoguchi T, Matsuo K, Tamada T, Suenaga Y. Secondary structure of human *de novo* evolved gene product NCYM analyzed by vacuum-ultraviolet circular dichroism. *Front Oncol* (2021) 11:688852. doi: 10.3389/fonc.2021.688852
- Van Oss SB, Carvunis A-R. *De novo* gene birth. *PLoS Genet* (2019) 15:e1008160. doi: 10.1371/journal.pgen.1008160
- Broeils LA, Ruiz-Orera J, Snel B, Hubner N, van Heesch S. Evolution and implications of *de novo* genes in humans. *Nat Ecol Evol* (2023) 7:804–15. doi: 10.1038/s41559-023-02014-y
- Weisman CM. The origins and functions of *de novo* genes: against all odds? *J Mol Evol* (2022) 90:244–57. doi: 10.1007/s00239-022-10055-3
- Bornberg-Bauer E, Hlouchova K, Lange A. Structure and function of naturally evolved *de novo* proteins. *Curr Opin Struct Biol* (2021) 68:175–83. doi: 10.1016/j.sbi.2020.11.010

## Conflict of interest

MH and KH are employees of NEC Solution Innovators. This does not alter the authors' adherence to all Frontiers in Oncology policies.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2023.1213678/full#supplementary-material>

- Heames B, Buchel F, Aube M, Tretyachenko V, Loginov D, Novak P, et al. Experimental characterization of *de novo* proteins and their unevolved random-sequence counterparts. *Nat Ecol Evol* (2023) 7:570–80. doi: 10.1038/s41559-023-02010-2
- Bungard D, Copple JS, Yan J, Chhun JJ, Kumirov VK, Foy SG, et al. Foldability of a natural *de novo* evolved protein. *Structure* (2017) 25:1687–1696.e4. doi: 10.1016/j.str.2017.09.006
- Lange A, Patel PH, Heames B, Damry AM, Saenger T, Jackson CJ, et al. Structural and functional characterization of a putative *de novo* gene in *Drosophila*. *Nat Commun* (2021) 12:1667. doi: 10.1038/s41467-021-21667-6
- Her C, Yeh Y, Krishnan VV. The ensemble of conformations of antifreeze glycoproteins (AFGP8): a study using nuclear magnetic resonance spectroscopy. *Biomolecules* (2019) 9:235. doi: 10.3390/biom9060235
- Armstrong BC, Krystal GW. Isolation and characterization of complementary DNA for N-cym, a gene encoded by the DNA strand opposite to N-myc. *Cell Growth Differ Mol Biol J Am Assoc Cancer Res* (1992) 3:385–90.
- Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* (2010) 5:e15004. doi: 10.1371/journal.pone.0015004
- Minagawa H, Onodera K, Fujita H, Sakamoto T, Akitomi J, Kaneko N, et al. Selection, characterization and application of artificial DNA aptamer containing appended bases with sub-nanomolar affinity for a salivary biomarker. *Sci Rep* (2017) 7:42716. doi: 10.1038/srep42716
- Cantu VA, Sadural J, Edwards R. PRINSEQ++, a multi-threaded tool for fast and efficient quality control and preprocessing of sequencing datasets. *PeerJ Prepr* (2019) 7:e27553v1. doi: 10.7287/peerj.preprints.27553v1
- Endo M, Katsuda Y, Hidaka K, Sugiyama H. Regulation of DNA methylation using different tensions of double strands constructed in a defined DNA nanostructure. *J Am Chem Soc* (2010) 132:1592–7. doi: 10.1021/ja907649w
- Rajendran A, Endo M, Hidaka K, Tran PLT, Mergny J-L, Sugiyama H. Controlling the stoichiometry and strand polarity of a tetramolecular G-quadruplex structure by using a DNA origami frame. *Nucleic Acids Res* (2013) 41:8738–47. doi: 10.1093/nar/gkt592
- Endo M, Inoue M, Suzuki Y, Masui C, Morinaga H, Hidaka K, et al. Regulation of B-Z conformational transition and complex formation with a Z-form binding protein by introduction of constraint to double-stranded DNA by using a DNA nanoscaffold. *Chem Eur J* (2013) 19:16887–90. doi: 10.1002/chem.201303830

24. Yang Y, Endo M, Suzuki Y, Hidaka K, Sugiyama H. Direct observation of the dual-switching behaviors corresponding to the state transition in a DNA nanoframe. *Chem Commun* (2014) 50:4211–3. doi: 10.1039/C4CC00489B
25. Rajendran A, Endo M, Hidaka K, Sugiyama H. Direct and single-molecule visualization of the solution-state structures of G-hairpin and G-triplex intermediates. *Angew Chem Int Ed Engl* (2014) 53:4107–12. doi: 10.1002/anie.201308903
26. Suzuki Y, Endo M, Cañas C, Ayora S, Alonso JC, Sugiyama H, et al. Direct analysis of Holliday junction resolving enzyme in a DNA origami nanostructure. *Nucleic Acids Res* (2014) 42:7421–8. doi: 10.1093/nar/gku320
27. Yamamoto S, De D, Hidaka K, Kim KK, Endo M, Sugiyama H. Single molecule visualization and characterization of Sox2-Pax6 complex formation on a regulatory DNA element using a DNA origami frame. *Nano Lett* (2014) 14:2286–92. doi: 10.1021/nl4044949
28. Endo M, Xing X, Zhou X, Emura T, Hidaka K, Tuesuwan B, et al. Single-molecule manipulation of the duplex formation and dissociation at the G-quadruplex/i-motif site in the DNA nanostructure. *ACS Nano* (2015) 9:9922–9. doi: 10.1021/acs.nano.5b03413
29. Ráz MH, Hidaka K, Sturla SJ, Sugiyama H, Endo M. Torsional constraints of DNA substrates impact cas9 cleavage. *J Am Chem Soc* (2016) 138:13842–5. doi: 10.1021/jacs.6b08915
30. Shimizu N, Yatabe K, Nagatani Y, Saijyo S, Kosuge T, Igarashi N. Software development for analysis of small-angle x-ray scattering data. *AIP Conf Proc* (2016) 1741:050017. doi: 10.1063/1.4952937
31. Savelyev A. Assessment of the DNA partial specific volume and hydration layer properties from CHARMM Drude polarizable and additive MD simulations. *Phys Chem Chem Phys* (2021) 23:10524–35. doi: 10.1039/d1cp00688f
32. Orthaber D, Bergmann A, Glatter O. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *J Appl Crystallogr* (2000) 33:218–25. doi: 10.1107/S0021889899015216
33. Svergun D, Koch MH, Timmins P, May RP. *Small angle X-ray and neutron scattering from solutions of biological macromolecules*. United Kingdom: Oxford University Press (2013).
34. Svergun DI, Petoukhov MV, Koch MH, Biophys J. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* (2001) 80:2946–53. doi: 10.1016/S0006-3495(01)76260-1
35. Franke D, Svergun DI. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* (2009) 42:342–6. doi: 10.1107/s0021889809000338
36. Petoukhov MV, Konarev PV, Kikhney AG, Svergun DI. ATSAS 2.1 – towards automated and web-supported small-angle scattering data analysis. *J Appl Crystallogr* (2007) 40:s223–8. doi: 10.1107/s0021889807002853
37. Volkov VV, Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* (2003) 36:860–4. doi: 10.1107/s0021889803000268
38. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* (1999) 76:2879–86. doi: 10.1016/S0006-3495(99)77443-6
39. Gabel F, Engilberge S, Schmitt E, Thureau A, Mechulam Y, Pérez J, et al. Medical contrast agents as promising tools for biomacromolecular SAXS experiments. *Acta Crystallogr Sect D Struct Biol* (2022) 78:1120–30. doi: 10.1107/S2059798322007392
40. Matsuo T, Kono F, Fujiwara S. Effects of the cardiomyopathy-causing E244D mutation of troponin T on the structures of cardiac thin filaments studied by small-angle X-ray scattering. *J Struct Biol* (2019) 205:196–205. doi: 10.1016/j.jsb.2018.12.005
41. Matsuo T, Arata T, Oda T, Fujiwara S. Difference in hydration structures between F-actin and myosin subfragment-1 detected by small-angle X-ray and neutron scattering. *Biophysics (Oxf)* (2013) 9:99–106. doi: 10.2142/biophysics.9.99
42. Porod G. General theory. In: Glatter O, Kratky O, editors. *Small angle X-ray scattering*. London: Plenum Press. (1982) p. 17–52.
43. Rambo RP, Tainer JA. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the porod-debye law. *Biopolymers* (2011) 95:559–71. doi: 10.1002/Bip.21638
44. Matsuo T. Usefulness of medium-angle X-ray scattering for structural characterization of flexible proteins studied by computer simulations. *Biochem Biophys Res Commun* (2020) 525:830–5. doi: 10.1016/j.bbrc.2020.02.150
45. Matsuo T, Arluisson V, Wien F, Peters J. Structural information on bacterial amyloid and amyloid-DNA complex obtained by small-angle neutron/X-ray scattering. In: Arluisson V, Wien F, Marcoleta A, editors. *Bacterial amyloids: methods and protocols, methods in molecular biology*. New Jersey: Springer, Humana Press. (2022) p. 95–107.
46. Harpaz Y, Gerstein M, Chothia C. Volume changes on protein folding. *Structure* (1994) 2:641–9. doi: 10.1016/S0969-2126(00)00065-4
47. Ko SH, Gallatin GM, Liddle JA. Nanomanufacturing with DNA origami: factors affecting the kinetics and yield of quantum dot binding. *Adv Funct Mater* (2012) 22:1015–23. doi: 10.1002/adfm.201102077
48. Huang J, Suma A, Cui M, Grundmeier G, Carnevale V, Zhang Y, et al. Arranging small molecules with subnanometer precision on DNA origami substrates for the single-molecule investigation of protein–ligand interactions. *Small Struct* (2020) 1:2000038. doi: 10.1002/ssstr.202000038
49. Rosier BJHM, Cremers GAO, Engelen W, Merckx M, Brunsveld L, de Greef TFA. Incorporation of native antibodies and Fc-fusion proteins on DNA nanostructures via a modular conjugation strategy. *Chem Commun* (2017) 53:7393–6. doi: 10.1039/C7CC04178K
50. Ngo TA, Dinh H, Nguyen TM, Liew FF, Nakata E, Morii T. Protein adaptors assemble functional proteins on DNA scaffolds. *Chem Commun* (2019) 55:12428–46. doi: 10.1039/C9CC04661E
51. Funke JJ, Dietz H. Placing molecules with Bohr radius resolution using DNA origami. *Nat Nanotechnol* (2016) 11:47–52. doi: 10.1038/nnano.2015.240
52. Voigt NV, Tørring T, Rotaru A, Jacobsen MF, Ravnsbæk JB, Subramani R, et al. Single-molecule chemical reactions on DNA origami. *Nat Nanotechnol* (2010) 5:200–3. doi: 10.1038/nnano.2010.5
53. Zhao S, Tian R, Wu J, Liu S, Wang Y, Wen M, et al. A DNA origami-based aptamer nanoarray for potent and reversible anticoagulation in hemodialysis. *Nat Commun* (2021) 12:358. doi: 10.1038/s41467-020-20638-7
54. Rinker S, Ke Y, Liu Y, Chhabra R, Yan H. Self-assembled DNA nanostructures for distance-dependent multivalent ligand-protein binding. *Nat Nanotechnol* (2008) 3:418–22. doi: 10.1038/nnano.2008.164
55. Chhabra R, Sharma J, Ke Y, Liu Y, Rinker S, Lindsay S, et al. Spatially addressable multiprotein nanoarrays templated by aptamer-tagged DNA nanoarchitectures. *J Am Chem Soc* (2007) 129:10304–5. doi: 10.1021/ja072410u
56. Mei Q, Johnson RH, Wei X, Su F, Liu Y, Kelbauskas L, et al. On-chip isotachopheresis separation of functional DNA origami capture nanoarrays from cell lysate. *Nano Res* (2013) 6:712–9. doi: 10.1007/s12274-013-0347-1
57. Zhou Z, Song X, Berezov A, Zhang G, Li Y, Zhang H, et al. Human glucocorticoid-induced TNF receptor ligand regulates its signaling activity through multiple oligomerization states. *Proc Natl Acad Sci USA* (2008) 105:5465–70. doi: 10.1073/pnas.0711350105
58. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proc Natl Acad Sci USA* (2000) 97:8868–73. doi: 10.1073/pnas.160259697