

# NAOMEE: Nucleic Acid Origami Minimal Exchange Format

A large curated corpus of Nucleic Acid Origami provided in a minimal format aided by a Domain Specific Language

Jordan Connolly, Ben Shirt-Ediss, Emanuela Torelli, Leanne Hobbs, Jaume Bacardit and Natalio Krasnogor

Join NAOMEE's Jamboard by scanning this 

We seek your feedback about how, what, & when to markup DNA, RNA or DNA/RNA origami, and ideas on how to best engage the community.



Please add your ideas and contact details (optional)

## Nucleic Acid Origami

Nucleic Acid origami is a technique that uses long strands of nucleotide bases and a set of oligonucleotide staples in a mixture to form DNA / RNA nanostructures.

## Aim

To allow exchange of Nucleic Acid origami experimental data in a minimal format, containing all useful details to aid lab and computational experimental replication.

Origami literature contains designs of novel structures or structures modified and applied to new domains.

Certain details regarding design files, experimental protocol and important metadata is often not indicated.

We provide ideas towards a minimum information standard for both Nucleic Acid Origami Experiments and Designs.

## COLLECT LITERATURE

- Gathered 1903 potentially relevant Literature Papers
- Extracted Paper PDF and Supporting Material

## FILTER PAPERS

- Use regular expressions to tag paper content
- Manually curated over 900 relevant papers that contain DNA/RNA Origami instances and details

## MINIMAL ORIGAMI REPRESENTATION

### Structure and Shape Detail

- Structure size (nm)
- Dimensions (2D, 3D)
- Scaffold Name
- Scaffold Length (bases)

### Lab Protocol Detail

- Temperature Ramp
- Buffer Contents
- Scaffold / Staple Molarity

### Staple Derived Detail

- Number of Staples
- Staple Content Statistics

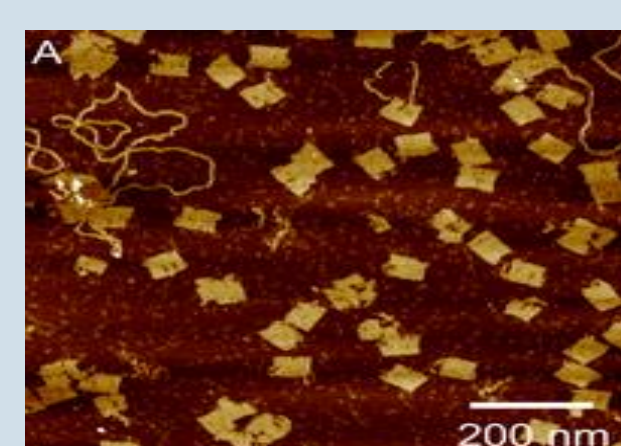
### Characterisation Detail

- Yield (%)
- Characterisation method

## Sequences

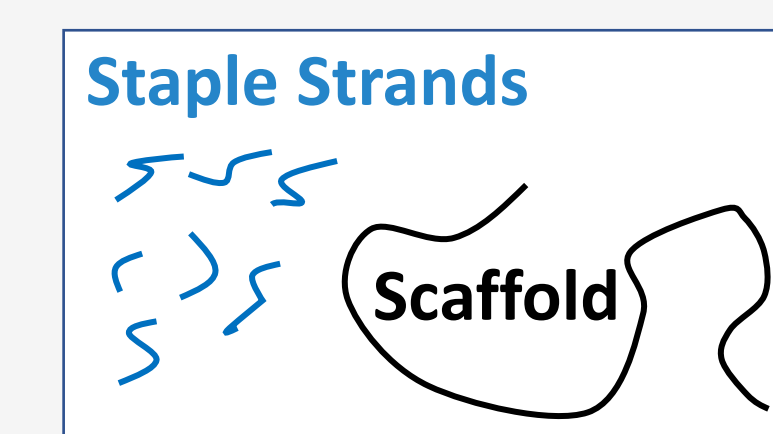
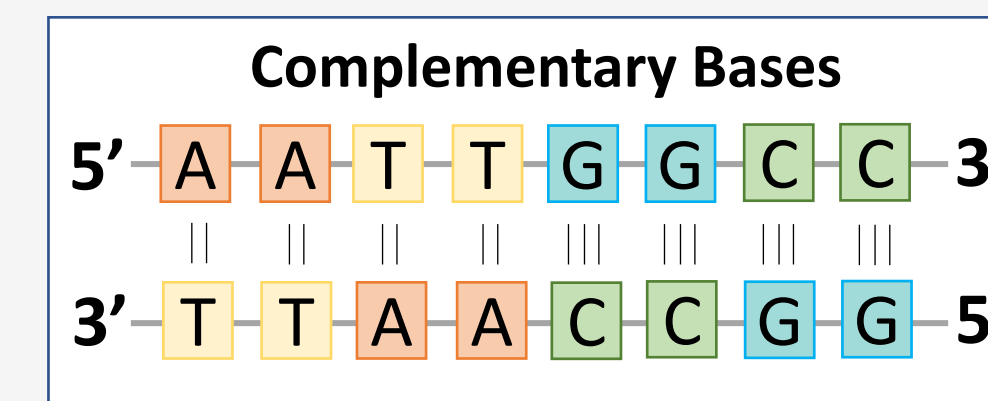
- Staple Sequences
- Scaffold Sequences
- Modified Sequences

## Images

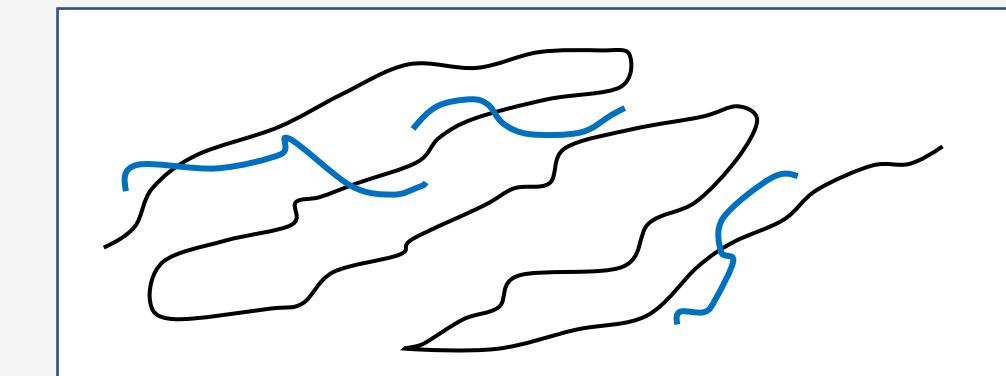


## Markdown Language

- Step 1: Input Set of Sequences
- Step 2: Mark Position of Modification
- Step 3: Label Modification Type
- Step 4: Add Description of Staple



Scaffold Strand is routed



Staple strands bind to scaffold, pinning it.

## STAPLE TO SCAFFOLD MAPPING

- Column 1) ID
- Column 2) Sequence
- Column 3) Sections
- Column 4) Section lengths
- Column 5) Scaffold domains respective staple sections hybridise to
- Column 6) Scaffold bases respective staple bases hybridise to

Staple	Sequence	Sections	Section lengths	Scaffold domains	Staple bases
Staple 1	ACTGACTGA	3	8;16;8	91;94;102	1212;1211;1210;1209
Staple 2	TGACGTAGT	2	8;40	95;90	1286;1285;1284;1283
Staple 3	ACTGAGTAC	3	24;16;8	89;84;80	1164;1163;1162;1161
Staple 4	ACGTAGTAC	2	8;37	85;88	1074;1073;1072;1071

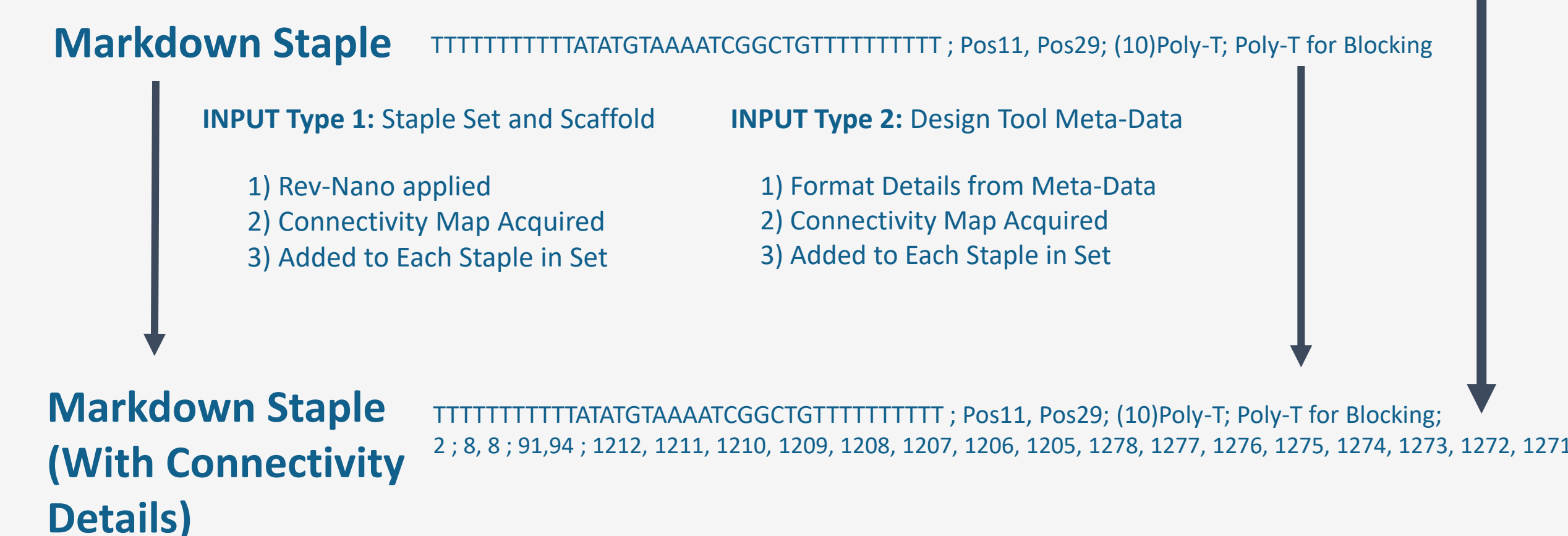
Staple	Sequence	Sections	Section lengths	Scaffold domains	Staple bases
Staple 1	ACTGACTGA	3	8;16;8	91;94;102	1212;1211;1210;1209
Staple 2	TGACGTAGT	2	8;40	95;90	1286;1285;1284;1283
Staple 3	ACTGAGTAC	3	24;16;8	89;84;80	1164;1163;1162;1161
Staple 4	ACGTAGTAC	2	8;37	85;88	1074;1073;1072;1071

## DESIGNS WHERE STAPLES ARE MAPPED TO SCAFFOLD ARE CONNECTIVITY MAPS

Format	Modified Staple
As presented in Literature	15[2] 18[2] TTTTTTTTTTATATGTAATAATCGGCTGTTTTTTTTT 38 Poly-Ts
Raw Sequence (Without formatting)	TTTTTTTTTTATATGTAATAATCGGCTGTTTTTTTTT
Modified Positions Marked	TTTTTTTTTTATATGTAATAATCGGCTGTTTTTTTTT ; Pos11, Pos29;
Type of Modification Added	TTTTTTTTTTATATGTAATAATCGGCTGTTTTTTTTT ; Pos11, Pos29; (10)Poly-T
Description of Modification	TTTTTTTTTTATATGTAATAATCGGCTGTTTTTTTTT ; Pos11, Pos29; (10)Poly-T; Poly-T for Blocking

## Connectivity Map for Nucleic Acid Origami

Step 5: Use Design Tool Meta-data to add Connectivity Map or infer from raw staple and scaffold sequences using the RevNano Tool.



## Minimum Exchange Format Applications

- Format provides easier sharing of origami designs and experimental protocols.
- Improved standardisation for use in downstream computational tools.

## Why?

The value added by this work would be to allow researchers working on data derived from lab experiments quickly and efficiently share raw or detailed sequences and instances of origami nanostructures with sufficient detail for reproduction of lab results and for use in downstream computational tools such as simulation, sequence optimisation, machine learning and more.