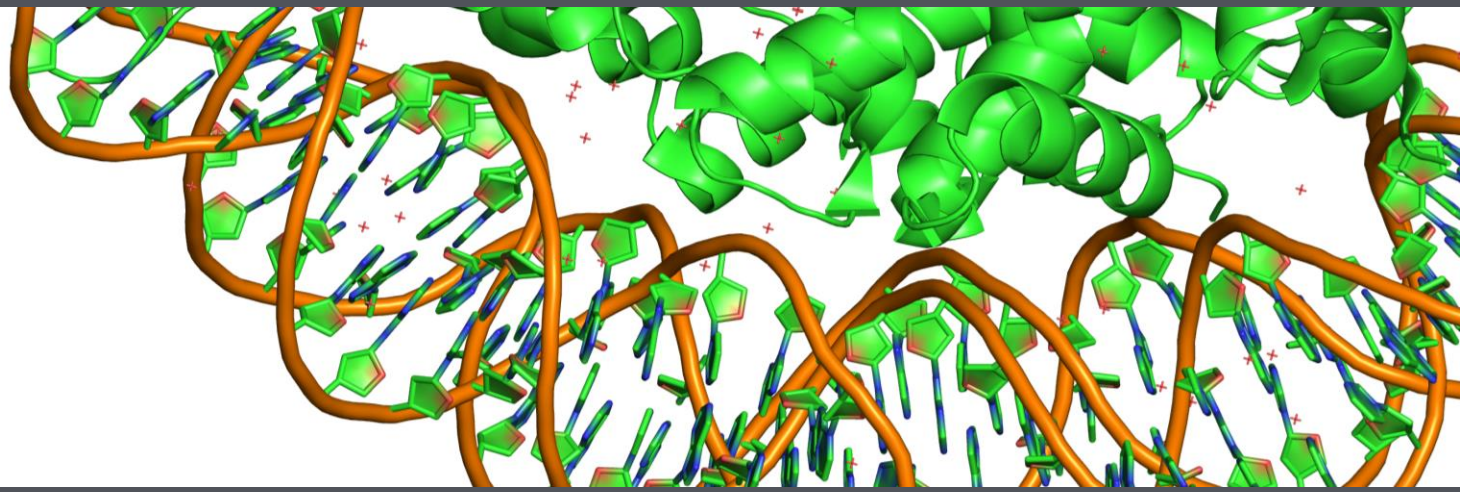
Reading School of Pharmacy

University of Reading

CRYSTALLOGRAPHY OF NUCLEIC ACIDS



Dr James Hall

James.hall@reading.ac.uk

CCP4 Workshop 2022

Copyright University of Reading

OUTLINE

University of Reading

1) Introduction to nucleic acid structure

2) Comparison with protein crystallography - similarities, differences and opportunities

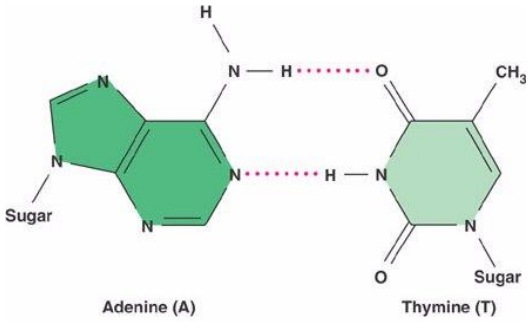
3) Useful tools…

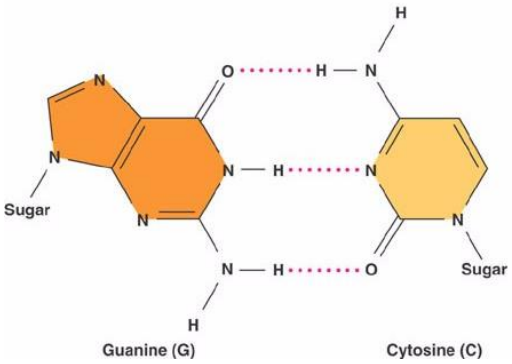
4) Questions?

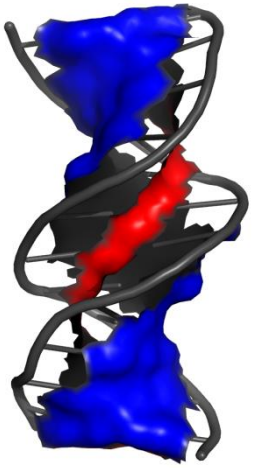
NUCLEIC ACID STRUCTURE

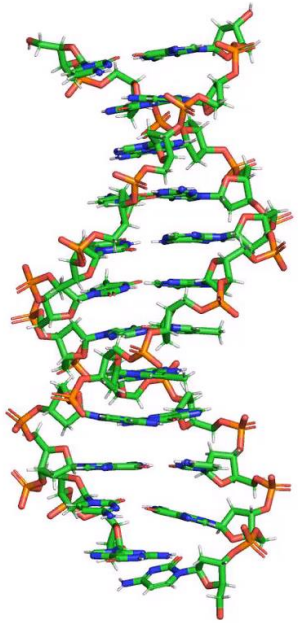
University of Reading

- Complementary base pairs form DNA ladder:

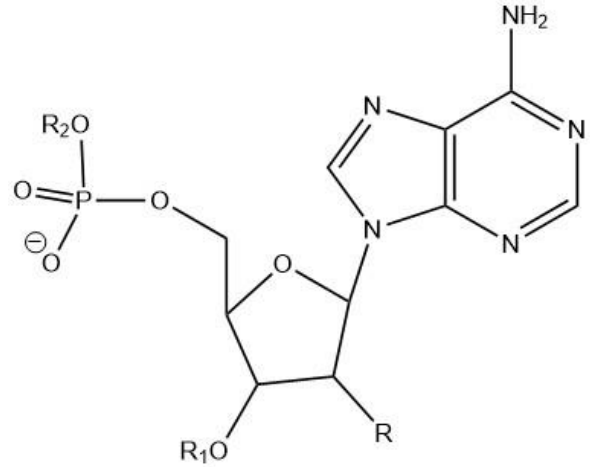








- Sequence written as base code: GCTTAAGAGGGCTCTGAGGA…..



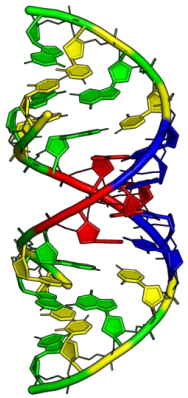
DNA – R = H

RNA – R = OH

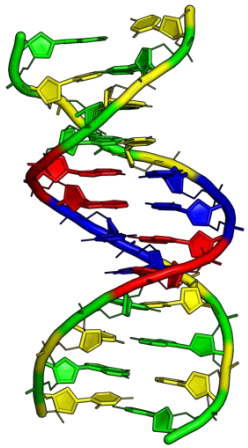
NUCLEIC ACID STRUCTURE

University of Reading

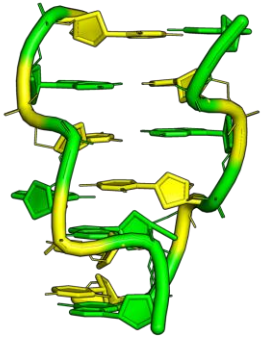
- However, nucleic acids are far more complex than this…..



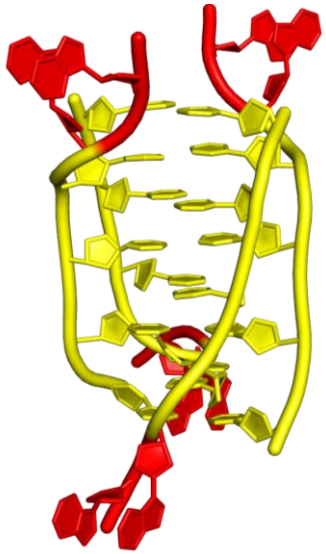
A-DNA
(Low humidity)



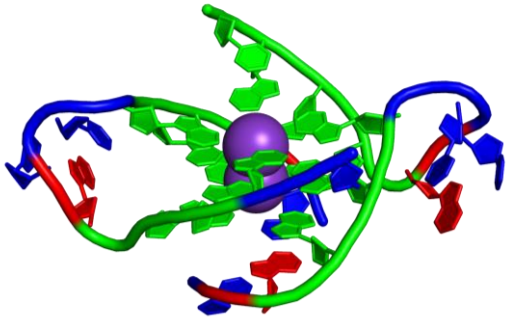
B-DNA
(Sequence)



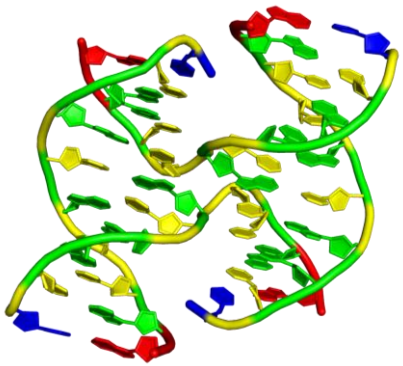
Z-DNA
(Left-handed, GC-repeats)



I-motif
(C rich, H^+ preferred)



Quadruplex
(G rich, K^+ preferred)

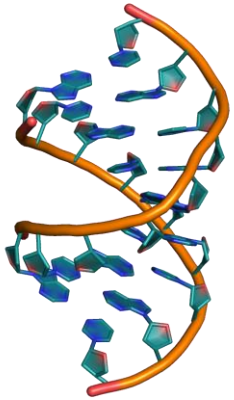


Holliday Junction
(Group 2 metals, sequence)

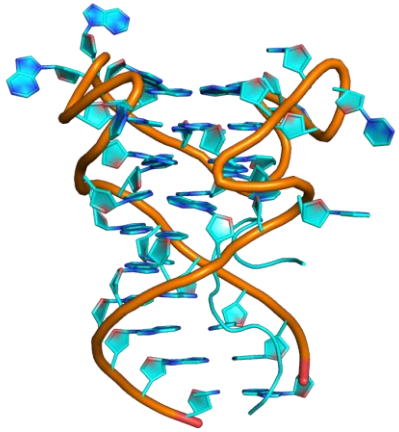
NUCLEIC ACID STRUCTURE

University of Reading

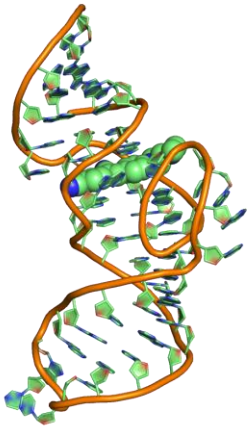
- RNA can be far more complex than DNA…..



A-RNA

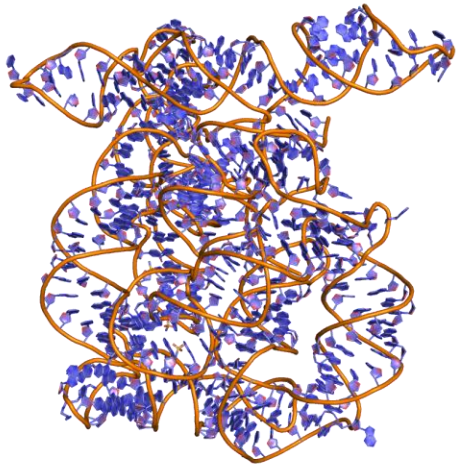


RNA Quadruplex

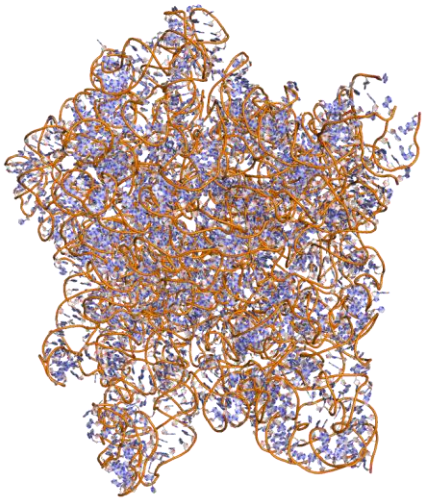


Aptamer

tRNA



Ribosyme



Ribosome subunit

STRUCTURAL VARIATION

University of Reading

- What causes this structural variation?

- Sequence – yes and no….

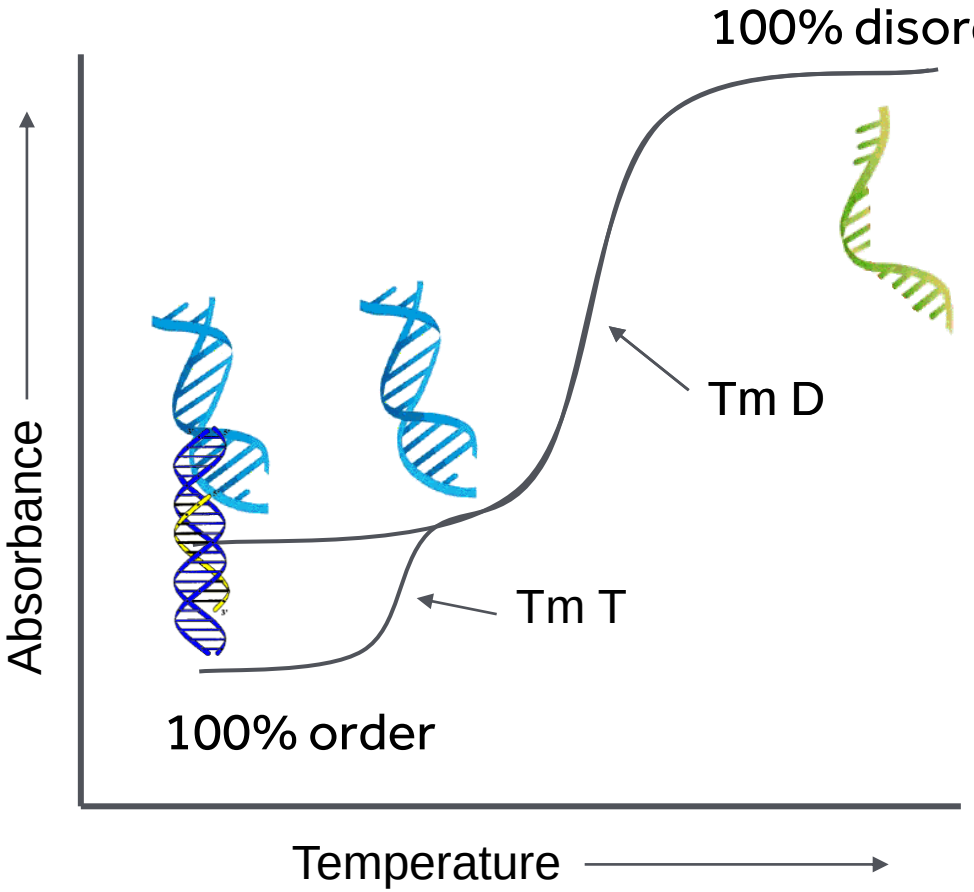
- Environment – yes and no…..

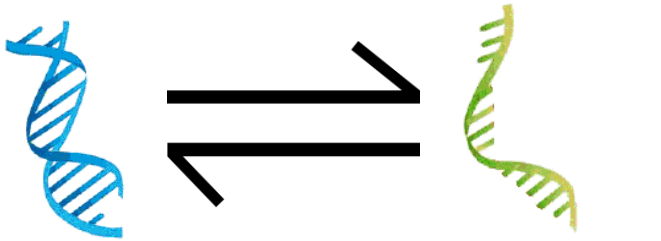
- Answer – both play a role. Understanding this can help with obtaining hits…

- **Nucleic acid structure is less dependent on sequence than proteins**

NUCLEIC ACID STRUCTURE

- Nucleic acid structures often exist in equilibrium….understanding this can help with crystallization





Large sequences – regions of disorder

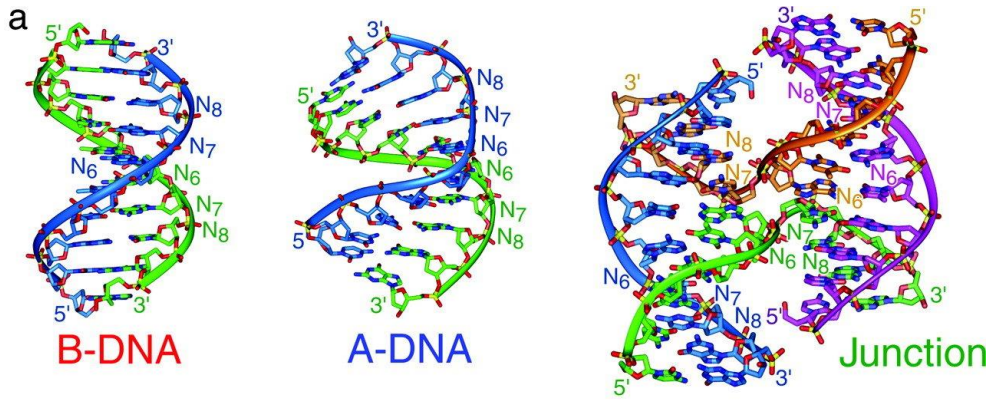
Small sequences – Ordered or denatured

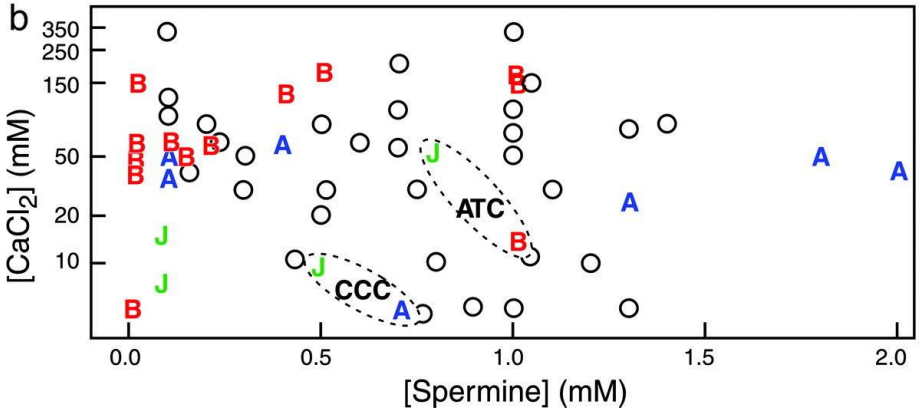
Melting temperature - T_m = 50% order/disorder

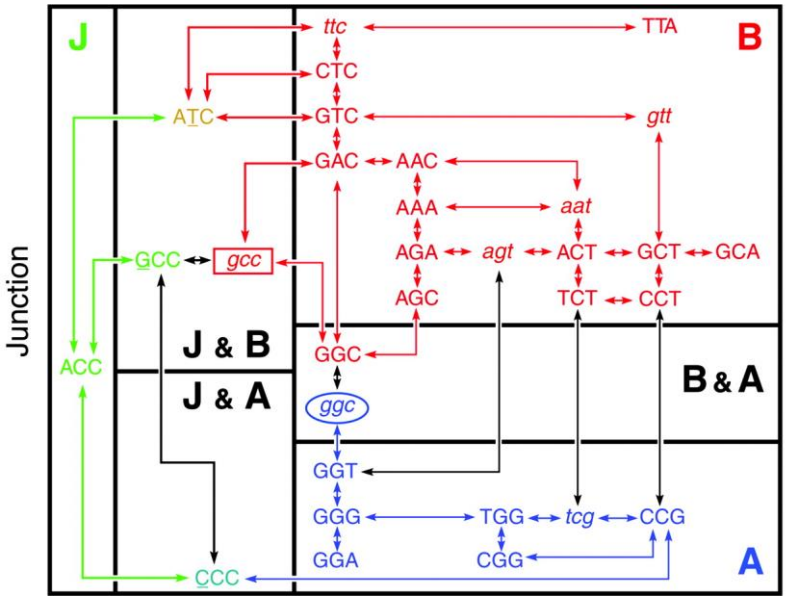
NUCLEIC ACID STRUCTURE

University of Reading

- Structure of small nucleic acids (<50 bases) is highly dependent on environment and sequence….challenging when it comes to structure solution!







d(CCnnn$N_6N_7N_8$GG) e.g. ATC = CCGATATCGG

Hayes, F.A. *et al. Proc Natl. Acad. Sci. USA.* 102 (20), 7157-7162, **2005**.

CHARACTERISATION

- UV/visible spectroscopy – characterise the stability of your system in your conditions
- Will give you T_m, not structure

- Easy to do:

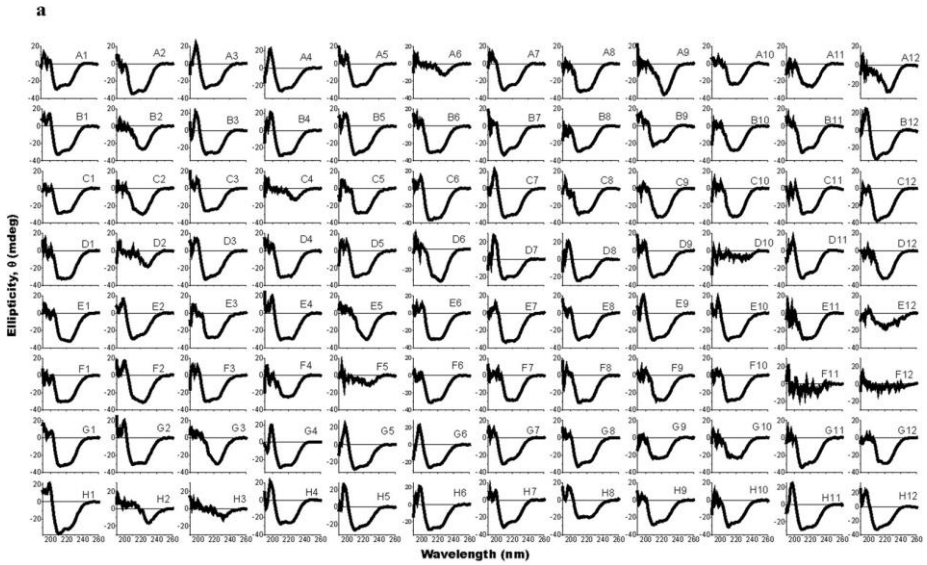
1. Add enough nucleic acid to a cuvette so absorption 0.1-1
2. Measure UV spectrum at lowest temperature (5-20°c)
3. Increase temperature 1-5°c
4. Measure again
5. Repeat until 90°c

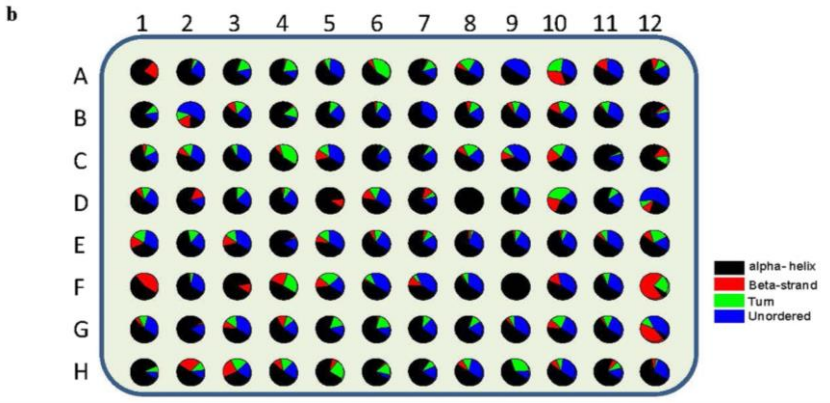
Will take approx. 2h. May save you months of failed experiments and help to identify crystallization temperature

CHARACTERISATION

- Circular dichroism – characterise the structure of your system in crystallization conditions
- Will give you structural information (very low resolution!)

Will take approx. 8h on B23 using multi-well plate (longer at home). May save you months of failed experiments

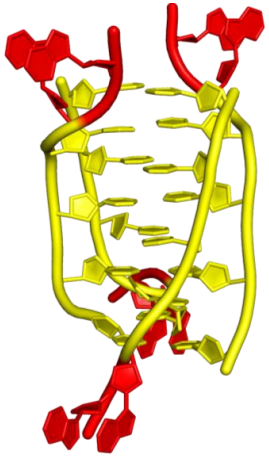




Hussain, R. *et al. Scientific Reports.* 6, 38028, **2016**.

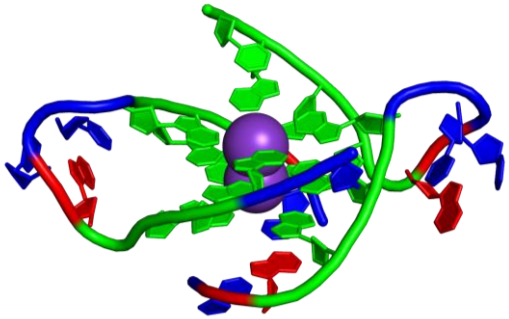
CHARACTERISATION

University of Reading



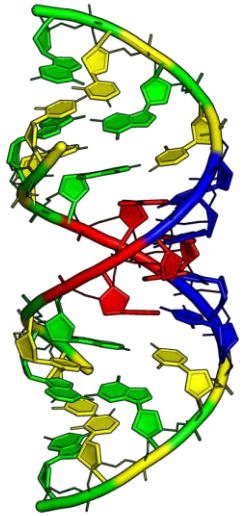
I-motif
(C rich, H^+ preferred)

Ions/salts far less important



Quadruplex
(G rich, K^+ preferred)

pH far less important



A-DNA
(Low humidity)

pH and ions/salts far less important

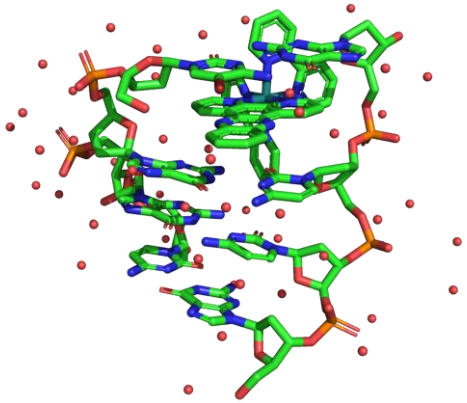
Knowledge of sequence behaviour/preferences is essential prior to crystallization

COMPARISON WITH PROTEIN CRYSTALLOGRAPHY

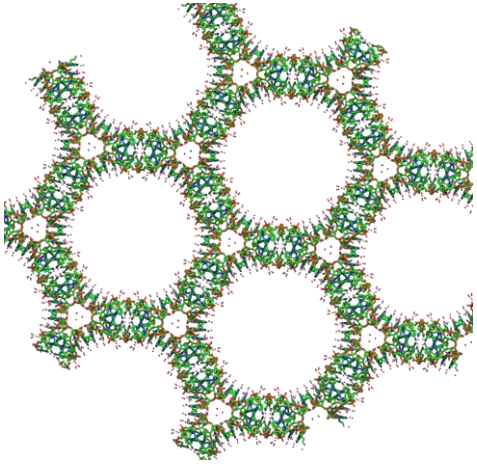
University of Reading

Length of sequence

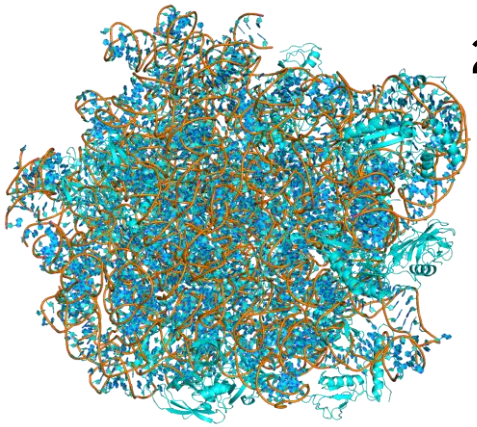
- A few bases…..(1-3 kDa..ca. 300 Da per base, DNA or RNA)



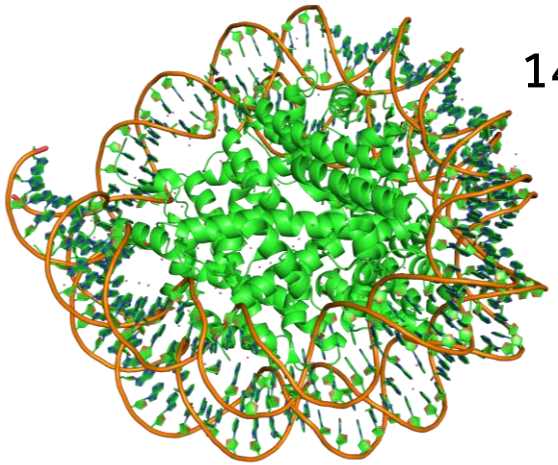
2x 4 base strands
1x Ligand
Waters + cations
5LFW
$P6_4\ 2\ 2$
67.34, 67.34, 35.79



Large extended structures (RNA, DNA origami or DNA/RNA-Protein)



2923 bases RNA
1.36 MDa
4WCE
$P6_5\ 2\ 2_1$
279.76, 279.76, 872.73



146 base pairs DNA
198.9 kDa
1P3K
$P2_1\ 2_1\ 2_1$
104.99, 109.68, 180.72

COMPARISON WITH PROTEIN CRYSTALLOGRAPHY

- Crystallization methods – very similar!
- Sitting drop/hanging drop vapour diffusion is commonly used
- Nucleic acid-specific screens:

Hampton research – **Natrix**

Molecular Dimensions – Helix, MIDASplus

Salt and precipitant concentration are the most important to screen

pH less so unless system is pH responsive

nucleic acid structural stability is VERY sensitive to concentration and type of salt, as nucleic acids are polyanionic

2-methyl-2,4-pentanediol (MPD, Hexylene glycol) precipitant of choice, although this is system specific…MPD is a cryo protectant

Anecdotally – small nucleic acids don't seem to like PEG….

COMPARISON WITH PROTEIN CRYSTALLOGRAPHY

University of Reading

- Data collection – also very similar to protein crystallography, except:

Nucleic acids are far more resistant to radiation damage than proteins

Outcome of using model protein-DNA complex over a dose range 2.07-44.63 MGy

"At low doses the protein was observed to be susceptible to radiation damage while the DNA was far more resistant, damage only being observed at significantly higher doses."

For nucleic acid-only crystals, significantly higher dose can be used with less risk of damage.....

Site-specific damage manifests primarily to the phosphate groups

Bury, C. *et al. J. Synchrotron. Rad.* 22, 213-224, **2015**.

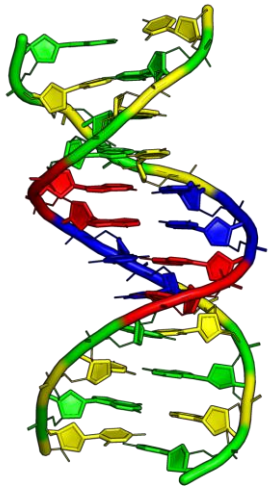
Bugris, V. *et al. J. Synchrotron Rad.* 26, 998-1009, **2019**.

COMPARISON WITH PROTEIN CRYSTALLOGRAPHY

University of Reading

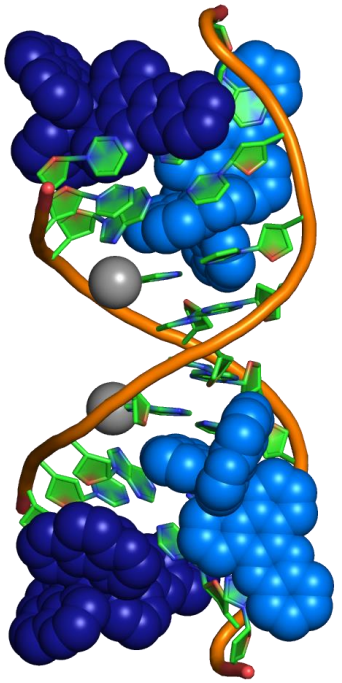
- Structure solution – same options (MAD/SAD, MR……etc), different challenges

- For small nucleic acids, molecular replacement can be challenging, particularly if ligands are involved:



B-DNA

Add ligand….

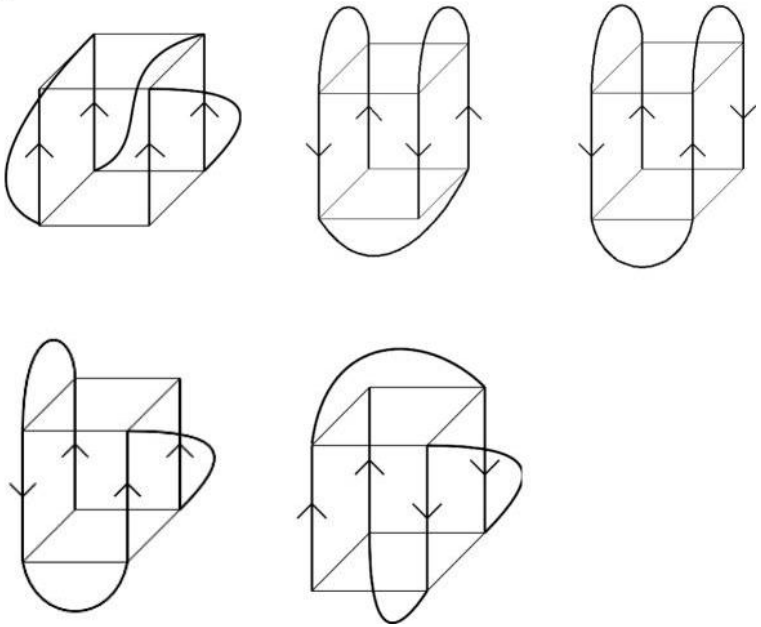


Hall, J. P. *et al. Proc Natl Acad Sci USA.* **108,** 17610-17617 (2011).

COMPARISON WITH PROTEIN CRYSTALLOGRAPHY

University of Reading

- Structure solution – same options (MAD/SAD, MR……etc), different challenges
- Flexibility in DNA and RNA is in the phosphate backbone…..
- Base pairs are less flexible but also have less scattering than phosphate groups
- Topological flexibility of sequence also a serious issue:



G-rich RNA and DNA can form quadruplexes

Left are five possible topologies…..there are over 20 theoretical possibilities, not all of which have been structurally characterised

Burge, S. *et al. Nucleic Acids Res.* 34 (19), 5402-5415, **2006**.

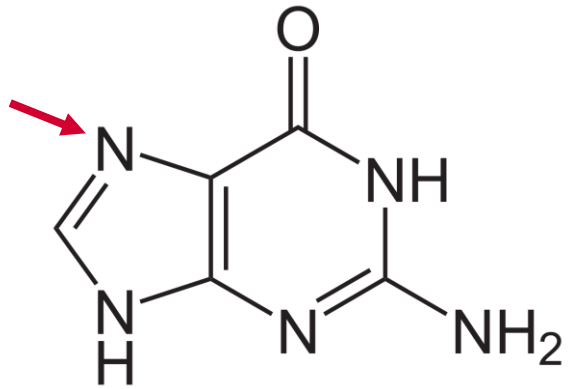
COMPARISON WITH PROTEIN CRYSTALLOGRAPHY

- Opportunity – SAD/MAD (radiation resistant so multiple datasets rarely a problem)

- Large nucleic acids (NAs) expressed and extracted – difficult to incorporate modifications

- If your sequence contains guanine…..get soaking!

N7 – excellent group to coordinate metal centres

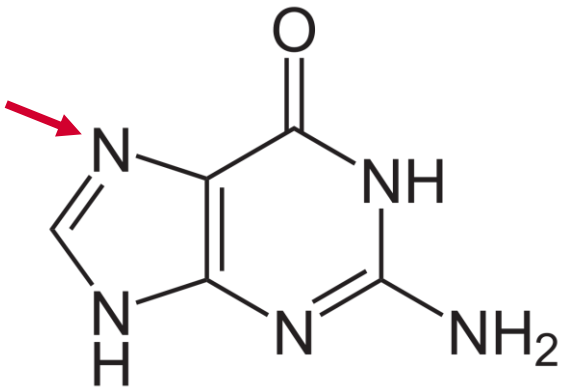


COMPARISON WITH PROTEIN CRYSTALLOGRAPHY

University of Reading

- Opportunity – SAD/MAD (radiation resistant so multiple datasets rarely a problem)
- If your sequence contains guanine…..get soaking!

N7 – excellent group to coordinate metal centres

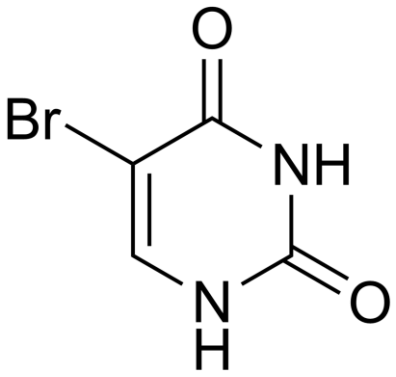


- Elements with atomic number higher than Ca^{2+} can work well
- Examples include:
- Sr^{2+}, Ba^{2+} (this is excellent), Mn^{2+}, Co^{2+}, Ni^{2+}, Zn^{2+} - avoid Fe^{2+} and Cu^{2+} *Cis*-platin will bind at N7 but low solubility can be challenging
- Natrix screen already contains some of the above ions for this reason!

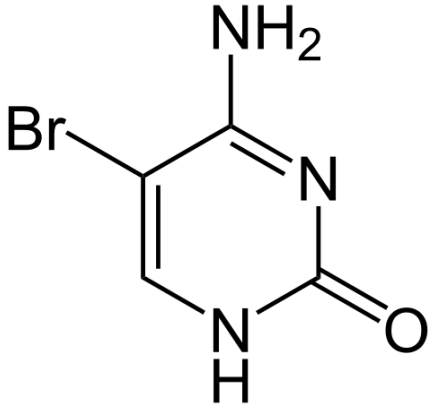
COMPARISON WITH PROTEIN CRYSTALLOGRAPHY

- Opportunity – SAD/MAD (radiation resistant so multiple datasets rarely a problem)
- Large nucleic acids (NAs) expressed and extracted – difficult to incorporate modifications

- Small NAs (up to 100 bases) are normally purchased:
- Merck, Eurogentec, ATDBio – all good suppliers (HPLC purification essential!)



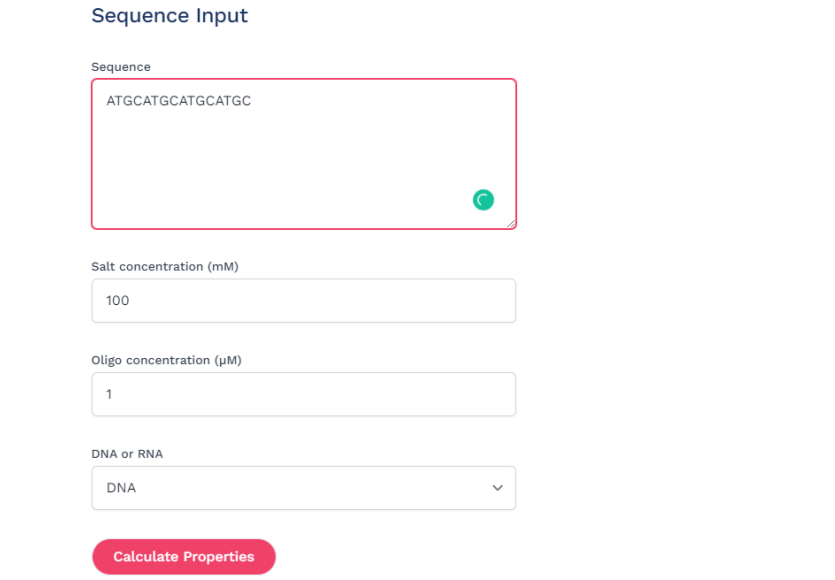
5-Br-U (replaces U/T)

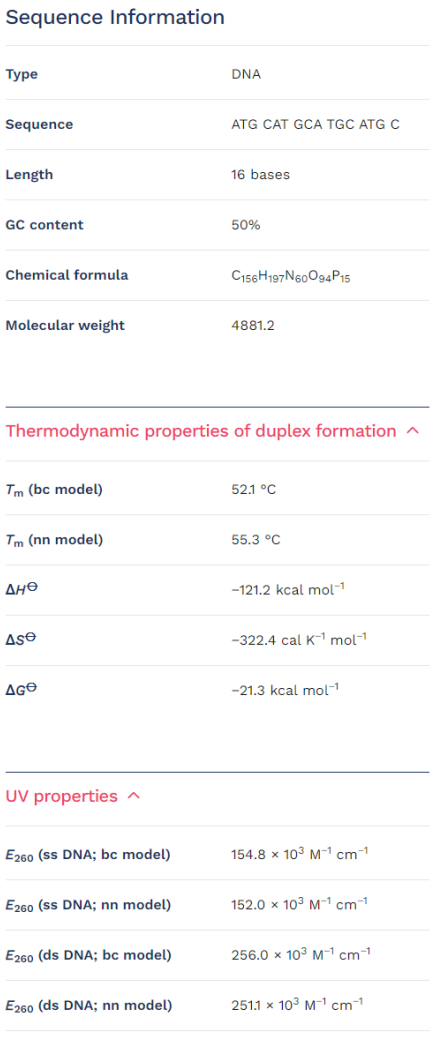


5-Br-C (replaces C)

USEFUL TOOLS - ATDBIO

- https://atdbio.com/tools/oligo-calculator - calculate expected stability of duplex (underestimates for small oligos), mass, sequence composition etc



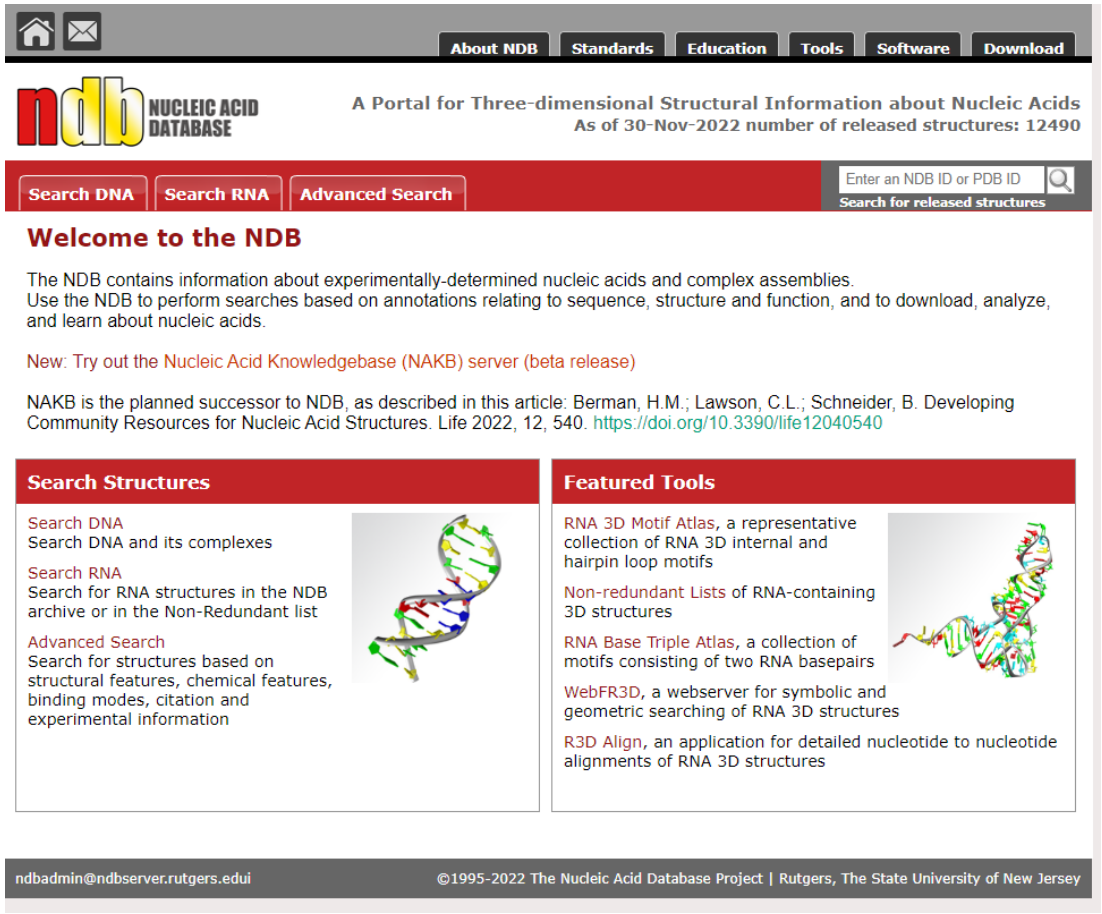


USEFUL TOOLS - NDB

- Nucleic Acid Database - http://ndbserver.rutgers.edu/

Soon to be replaced by Nucleic Acid Knowledgebase - https://beta.nakb.org/

The PDB for nucleic acids



USEFUL TOOLS - NDB

University of Reading

NUCLEIC ACID DATABASE

A Portal for Three-dimensional Structural Information about Nucleic Acids
As of 30-Nov-2022 number of released structures: 12490

Search DNA | Search RNA | Advanced Search | Enter an NDB ID or PDB ID | Search for released structures

DNA Search Options:

Polymer
- (•) All
- ( ) DNA Only
- ( ) Protein DNA Complexes
- ( ) Drug DNA Complexes
- ( ) Hybrids and Chimera
- ( ) Peptide Nucleic Acid / Mimetics

Protein Function
- (•) All
- ( ) Enzymes
- ( ) Structural
- ( ) Regulatory

Structural Features
- (•) All
- ( ) Single Stranded
- ( ) A DNA
- ( ) B DNA
- ( ) Z DNA
- ( ) Other Double Helical Structures
- ( ) Triple helices
- ( ) Quadruple helices

Experimental Method
- (•) All
- ( ) XRAY
- ( ) NMR

Text Search

Filter results by text search

Use this option to narrow your results down considerably (>50% reduction) using any word seen in the results page. Eg: Any author name found in the right side

Polymer Type: All + Protein Function: All + Structural Features: All + Experimental Method: All
Results: 8575 Download results as an excel file Gallery view

NDB ID: 7VRU PDB ID: 7VRU Release: 2022-11-30		Title: Crystal structure of PacII_M1M2S-DNA-SAH complex Classification: TRANSFERASE/DNA Authors: Zhu, J., Gao, Y., Wang, Y., Zhan, Q., Feng, H., Luo, X., Li, P., Liu, S., Hou, H., Gao, P. Citation: Molecular insights into DNA recognition and methylation by non-canonical type I restriction-modification systems. Nat Commun 13 pp.6391 - 6391 2022 Experiment: X-RAY DIFFRACTION Resolution:2.4Å R work:0.189 R free:0.234
NDB ID: 7VS4 PDB ID: 7VS4 Release: 2022-11-30		Title: Crystal structure of PacII_M1M2S-DNA(m6A)-SAH complex Classification: TRANSFERASE/DNA Authors: Zhu, J., Gao, Y., Wang, Y., Zhan, Q., Feng, H., Luo, X., Li, P., Liu, S., Hou, H., Gao, P. Citation: Molecular insights into DNA recognition and methylation by non-canonical type I restriction-modification systems. Nat Commun 13 pp.6391 - 6391 2022 Experiment: X-RAY DIFFRACTION Resolution:2.55Å R work:0.18 R free:0.236
NDB ID: 7ZKL PDB ID: 7ZKL Release: 2022-11-30		Title: X-ray structure of the complex between human alpha thrombin and a pseudo-cyclic thrombin binding aptamer (TBA-NNp/DDp) - Crystal form alpha Classification: HYDROLASE Authors: Troisi, R., Riccardi, C., Perez de Carvasal, K., Smietana, M., Morvan, F., Del Vecchio, P., Montesarchio, D., Sica, F. Citation: A terminal functionalization strategy reveals unusual binding abilities of anti-thrombin anticoagulant aptamers To Be Published pp. - 0 Experiment: X-RAY DIFFRACTION Resolution:3.18Å R work:0.19 R free:0.231
NDB ID: 7ZKM PDB ID: 7ZKM Release: 2022-11-30		Title: X-ray structure of the complex between human alpha thrombin and a pseudo-cyclic thrombin binding aptamer (TBA-NNp/DDp) - Crystal form beta Classification: HYDROLASE Authors: Troisi, R., Riccardi, C., Perez de Carvasal, K., Smietana, M., Morvan, F., Del Vecchio, P., Montesarchio, D., Sica, F. Citation: A terminal functionalization strategy reveals unusual binding abilities of anti-thrombin anticoagulant aptamers To Be Published pp. - 0 Experiment: X-RAY DIFFRACTION Resolution:2.0Å R work:0.226 R free:0.252

Experimental Details

AND | Space Group: select one

Unit Cell Dimensions:

AND | Alpha | = | ° AND | Beta | = | ° AND | Gamma | = | °

AND | a | = | Å AND | b | = | Å AND | c | = | Å

AND | Resolution better than: | Å AND | R-factor better than:

Citation Information

NDB ID NA2326 PDB ID 4JRC AND | Author Berman, H

AND | Publication Year: 2013 AND | Released Since: yyyy-mm-dd

RNA 3D Interactions

AND | Base Pair Interaction: select base pair

AND | Base Phosphate Interaction: select base phosphate

AND | Base Stack Interaction: select base stack

Relative Frequency of various RNA 3D Interactions:

AND | Base Pair Interaction select base pair with relative frequency: >= 0.1

AND | Base Phosphate Interaction select base phosphate with relative frequency: >= 0.1

AND | Base Stack Interaction select base stack with relative frequency: >= 0.1

RNA 3D Motifs

AND | Internal Loop Motif: All, Sarcin-ricin, Kink-turn

AND | Hairpin Loop Motif: All, T-loop, GNRA

Representative RNA 3D Structure Sets

AND | Resolution cutoff: select resolution cutoff

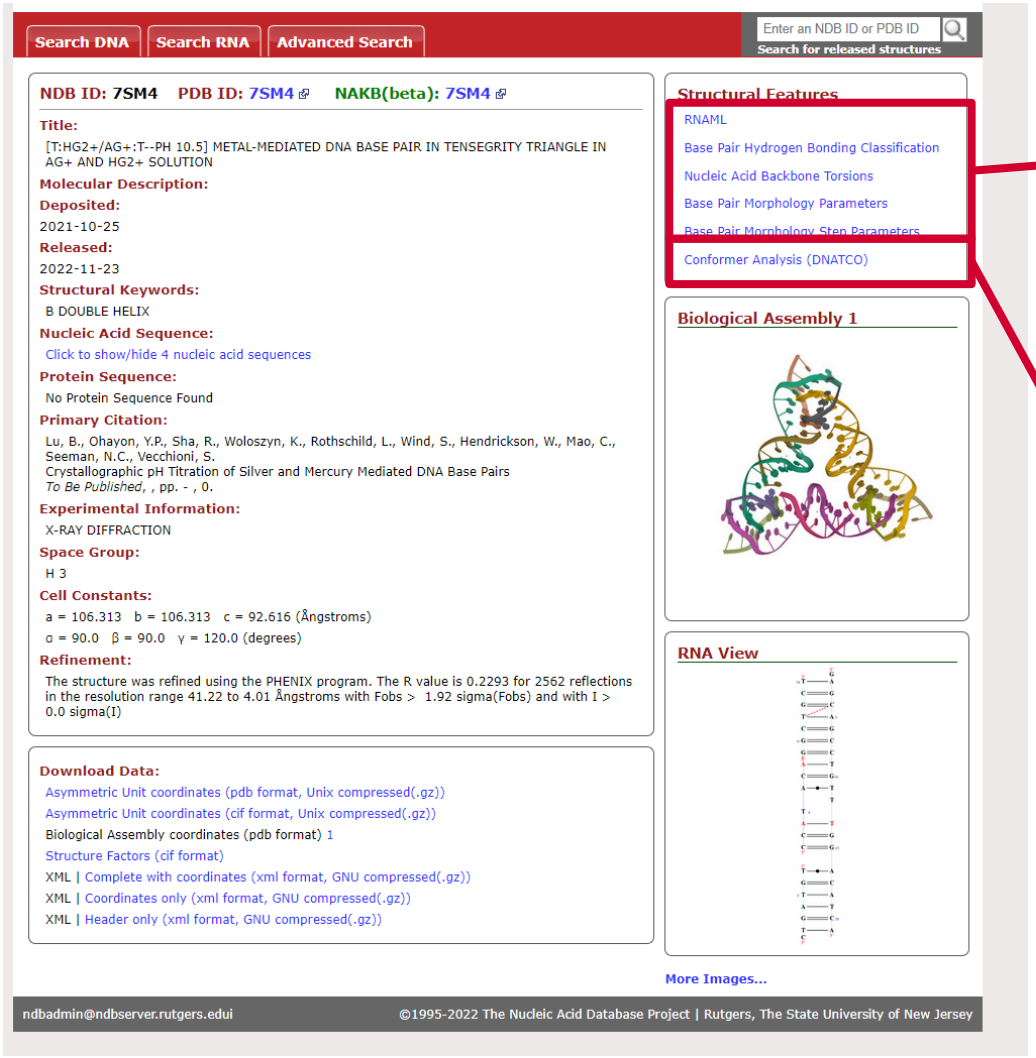
Sequence

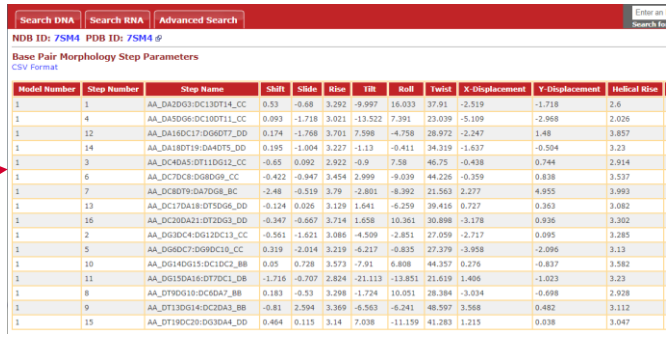
AND | Nucleic Acid Sequence Pattern:

AND | Oligonucleotide Sequence Length between and

USEFUL TOOLS - NDB

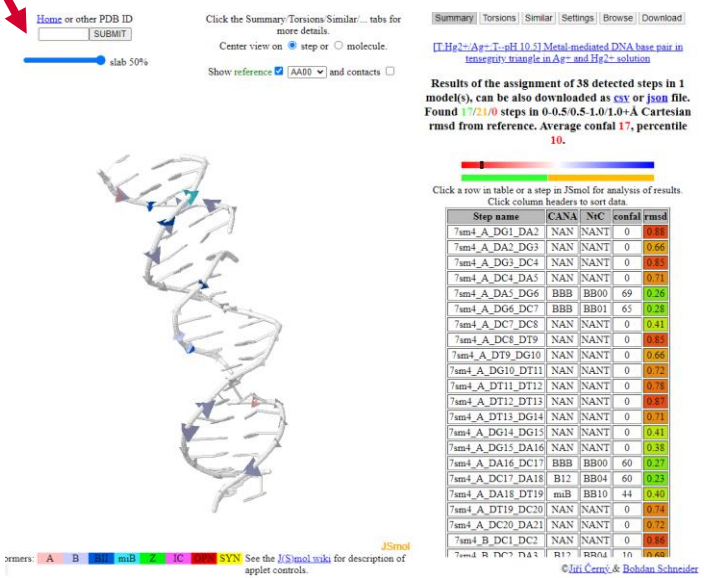
University of Reading





Model Number	Step Number	Step Name	Shift	Slide	Rise	Tilt	Roll	Twist	X-Displacement	Y-Displacement	Helical Rise
1	1	AA_DA2DG3:DC13DT14_CC	0.53	-0.68	3.292	-9.997	16.033	37.91	-2.519	-1.718	2.6
1	4	AA_DA5DG6:DC10DT11_CC	0.093	-1.718	3.021	-13.522	7.391	23.039	-5.109	-2.968	2.026
1	12	AA_DA16DC17:DG6DT7_DD	0.174	-1.768	3.701	7.598	-4.758	28.972	-2.247	1.48	3.857
1	14	AA_DA18DT19:DA4DT5_DD	0.195	-1.004	3.227	-1.13	-0.411	34.319	-1.637	-0.504	3.23
1	3	AA_DC4DA5:DT11DG12_CC	-0.65	0.092	2.922	-0.9	7.58	46.75	-0.438	0.744	2.914
1	6	AA_DC7DC8:DG8DG9_CC	-0.422	-0.947	3.454	2.999	-9.039	44.226	-0.359	0.838	3.537
1	7	AA_DC8DT9:DA7DG8_BC	-2.48	-0.519	3.79	-2.801	-8.392	21.563	2.277	4.955	3.993
1	13	AA_DC17DA18:DT5DG6_DD	-0.124	0.026	3.129	1.641	-6.259	39.416	0.727	0.363	3.082
1	16	AA_DC20DA21:DT2DG3_DD	-0.347	-0.667	3.714	1.658	10.361	30.898	-3.178	0.936	3.302
1	2	AA_DG3DC4:DG12DC13_CC	-0.561	-1.621	3.086	-4.509	-2.851	27.059	-2.717	0.095	3.285
1	5	AA_DG6DC7:DG9DC10_CC	0.319	-2.014	3.219	-6.217	-0.835	27.379	-3.958	-2.096	3.13
1	10	AA_DG14DG15:DC1DC2_BB	0.05	0.728	3.573	-7.91	6.808	44.357	0.276	-0.837	3.582
1	11	AA_DG15DA16:DT7DC1_DB	-1.716	-0.707	2.824	-21.113	-13.851	21.619	1.406	-1.023	3.23
1	8	AA_DT9DG10:DC6DA7_BB	0.183	-0.53	3.298	-1.724	10.051	28.384	-3.034	-0.698	2.928
1	9	AA_DT13DG14:DC2DA3_BB	-0.81	2.594	3.369	-6.563	-6.241	48.597	3.568	0.482	3.112
1	15	AA_DT19DC20:DG3DA4_DD	0.464	0.115	3.14	7.038	-11.159	41.283	1.215	0.038	3.047

http://web.x3dna.org/



Step name	CANA	NtC	confal	rmsd
7sm4_A_DG1_DA2	NAN	NANT	0	0.81
7sm4_A_DA2_DG3	NAN	NANT	0	0.66
7sm4_A_DG3_DC4	NAN	NANT	0	0.83
7sm4_A_DC4_DA5	NAN	NANT	0	0.71
7sm4_A_DA5_DG6	BBB	BB00	69	0.26
7sm4_A_DG6_DC7	BBB	BB01	65	0.28
7sm4_A_DC7_DC8	NAN	NANT	0	0.41
7sm4_A_DC8_DT9	NAN	NANT	0	0.85
7sm4_A_DT9_DG10	NAN	NANT	0	0.66
7sm4_A_DG10_DT11	NAN	NANT	0	0.72
7sm4_A_DT11_DT12	NAN	NANT	0	0.71
7sm4_A_DT12_DT13	NAN	NANT	0	0.87
7sm4_A_DT13_DG14	NAN	NANT	0	0.71
7sm4_A_DG14_DG15	NAN	NANT	0	0.41
7sm4_A_DG15_DA16	NAN	NANT	0	0.38
7sm4_A_DA16_DC17	BBB	BB00	60	0.27
7sm4_A_DC17_DA18	B12	BB04	60	0.23
7sm4_A_DA18_DT19	miB	BB10	44	0.40
7sm4_A_DT19_DC20	NAN	NANT	0	0.74
7sm4_A_DC20_DA21	NAN	NANT	0	0.72
7sm4_B_DC1_DC2	NAN	NANT	0	0.84

https://dnatco.datmos.org/

USEFUL TOOLS

Within CCP4 – model building of nucleic acids

- Arp/wARP (web service also available)
- Builds AND refines using electron density
- Relatively slow (compared to Nautilus) but refines, builds protein, solvent and can assign ligands

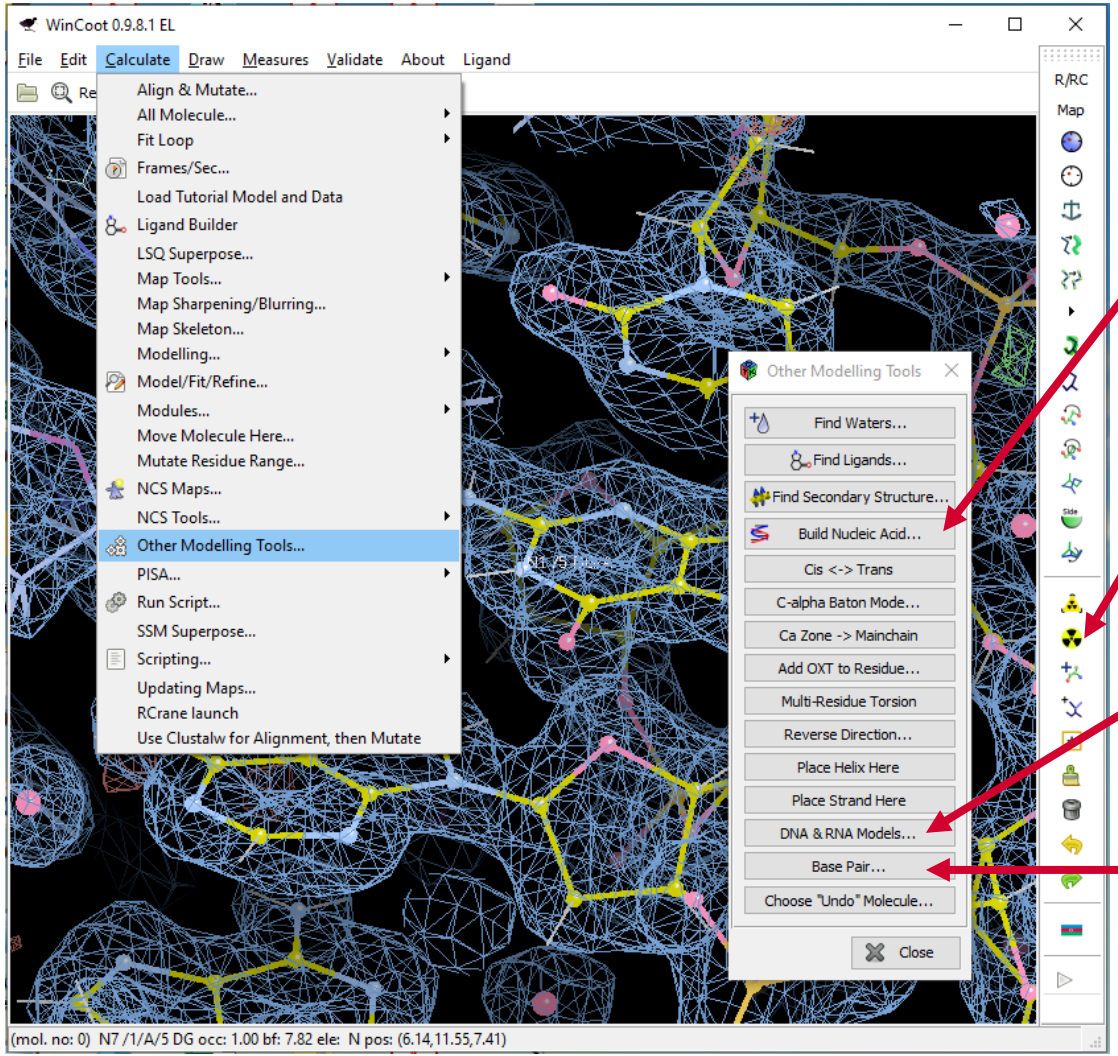
- Nautilus
- Builds from electron density
- VERY quick!

- Both require electron density and can use a starting model as a start point

USEFUL TOOLS

Within CCP4 – manual editing of nucleic acid structures

- Coot - I assume needs no introduction…



Try to build NA in density automatically

Substitute base type (mutate)

Generate idealised RNA/DNA strands for manual fitting

Generate the complementary base to complete a base pair

SUMMARY

University of Reading

- Understanding your nucleic acid system BEFORE crystallization can save you a lot of time and help you to select more likely growth conditions

- Careful soaking or incorporation of modified bases may help with structure solution – MR is very challenging with NAs and prediction of structure is far less certain

- Thanks to the developers for all their tools and guidance over the years – NA crystallography has become much more straightforward in the last decade thanks to their input and efforts!

- Any questions?