

- To reproduce the key results from Garg *et. al.*, we first need to find out what to model
- UniProt (<https://www.uniprot.org>) is a comprehensive biological sequence database
- A search for “thymidylate synthase” yields 50,099 results, including 566 that have been manually reviewed. This is much more than the 110 that were included by Garg *et. al.* in their 2015 paper!
- The two results for “thymidylate synthase wigglesworthia” is much more manageable. Select the result that has been reviewed.

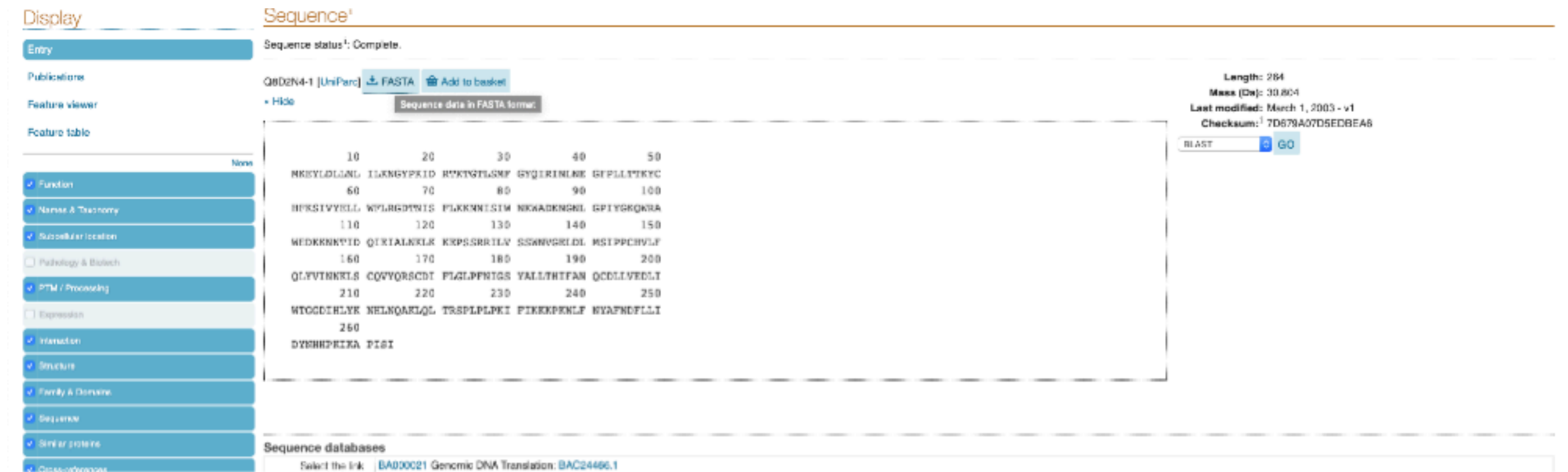
UniProt search results for "thymidylate synthase". The interface shows 50,099 results. The left sidebar includes filters for "Reviewed (566)" and "Unreviewed (49,533)". The main table lists several entries:

Entry	Entry name	Protein names	Gene names	Organism	Length
Q7W1A9	TYSY_BORPA	Thymidylate synthase	thyA BPB787	Bordetella pertussis (strain 12822 / ATCC BAA-687 / NCTC 13253)	323
C5BMA3	TYSY_TERTT	Thymidylate synthase	thyA TERTU_0365	Teredinibacter litoralis (strain ATCC 39867 / T7901)	277
Q8EV81	TYSY_MYCPE	Thymidylate synthase	thyA MYPE6570	Mycoplasma penetrans (strain HF-2)	289
A3PYA8	TYSY_MYCSJ	Thymidylate synthase	thyA Mjls_2089	Mycobacterium sp. (strain JLS)	266
POC6M5	TYSY_STAES	Thymidylate synthase	thyA SE_1120	Staphylococcus epidermidis (strain ATCC 12228)	318
Q6FZ31	TYSY_BAROU	Thymidylate synthase	thyA BQC8530	Bartonella quintana (strain Toulouse) (Rochalimaea quintana)	264
A6IWS2	TYSY_BART1	Thymidylate synthase	thyA BT_1568	Bartonella tribecorum (strain CIP 105476 / IBS 506)	264
Q6MID2	TYSY_BDEBA	Thymidylate synthase	thyA Bdc320	Bdellovibrio bacteriovorus (strain ATCC 15356 / DSM 50701 / NCIB 9529 / HD100)	264

UniProt search results for "thymidylate synthase wigglesworthia". The interface shows 2 results. The left sidebar includes filters for "Reviewed (1)" and "Unreviewed (1)". The main table lists two entries:

Entry	Entry name	Protein names	Gene names	Organism	Length
Q6Q2N4	TYSY_WIGBR	Thymidylate synthase	thyA WIGBR0300	Wigglesworthia glossinidia brevipalpis	264
H6Q4Z7	H6Q4Z7_WIGGL	Thymidylate synthase	thyA WIGMOR_0460	Wigglesworthia glossinidia endosymbiont of Glossina morsitans morsitans (Yale colony)	264

- To get the actual amino acid sequence, click on “Sequence” or scroll down and then click on FASTA. FASTA is a simple format for amino acid sequences based on one-letter codes.
- To run I-TASSER, you can visit the server interface at <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>, register, paste the sequence into the appropriate field, and press “Run I-TASSER”.
- *But* let’s spare their servers and avoid the wait by not all submitting the same job.

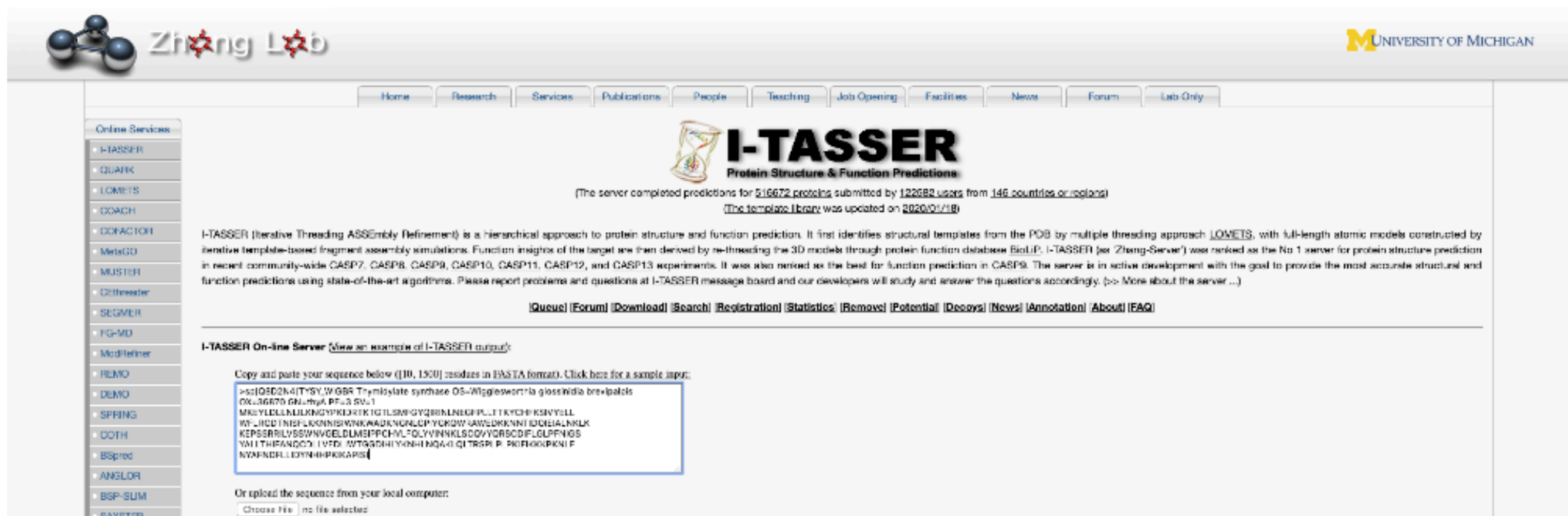


The screenshot shows the NCBI protein page for Q8D2N4-1. The left sidebar contains a 'Display' menu with options like 'Entry', 'Publications', 'Feature viewer', and 'Feature table'. The main content area shows the 'Sequence' tab with a FASTA format sequence. The sequence is as follows:

```

>Q8D2N4-1 [UniProt]
MKKYLQGLL ILKNGYKID RYKNTLSMF GYQIRINLE GFPLATKYC
60 70 80 90 100
IPKSIYVILL WYLRGPTWIS PLKKNKSTIM NKAADKNGRI GPIYKGNRA
110 120 130 140 150
MEDKKNYTD QRTALAKLK KPSRRRLV SSANVGRIDL MSIPPRVLF
160 170 180 190 200
QLVTHNKLK QVYVRSCT PLGLPPTGS YALLTHIFAN QCOLIVEDLI
210 220 230 240 250
WTGDDHLVK NELNQAQGL TRSLPLPKI PIKKPKNLF NYAPNDFLLI
260
DYSHPRKA PISI
  
```

Below the sequence, there are links to 'Sequence databases' and a 'Select the link' section.



The screenshot shows the I-TASSER Protein Structure & Function Predictions server interface. The top navigation bar includes links for Home, Research, Services, Publications, People, Teaching, Job Opening, Facilities, News, Forum, and Login Only. The main content area features the I-TASSER logo and a description of the server's capabilities. A text box on the left lists various online services, including I-TASSER, QUARK, LOMETS, COACH, DOFACON, MetaGO, MUSTER, COHSEDER, SEGMENT, PG-MD, ModRefiner, PRED, DEMO, SPHING, COPI, B3Psec, ANGLOR, BSP-SUM, and SAXTER. The central text box contains a sample input sequence in FASTA format:

```

>Q8D2N4-1 [UniProt]
MKKYLQGLL ILKNGYKID RYKNTLSMF GYQIRINLE GFPLATKYC
60 70 80 90 100
IPKSIYVILL WYLRGPTWIS PLKKNKSTIM NKAADKNGRI GPIYKGNRA
110 120 130 140 150
MEDKKNYTD QRTALAKLK KPSRRRLV SSANVGRIDL MSIPPRVLF
160 170 180 190 200
QLVTHNKLK QVYVRSCT PLGLPPTGS YALLTHIFAN QCOLIVEDLI
210 220 230 240 250
WTGDDHLVK NELNQAQGL TRSLPLPKI PIKKPKNLF NYAPNDFLLI
260
DYSHPRKA PISI
  
```

Below the sequence, there are links to 'Sequence databases' and a 'Select the link' section.