Journal of Undergraduate Chemistry Research, **2025**, 24 (3), 50

CLONING AND OVEREXPRESSION OF A THERMOPHILIC MALATE SYNTHASE

Jayde Bertoch*, Scott Adair*, Elisa Lago*, Elizabeth Pierce† and Bruce R. Howard†

Department of Chemistry and Physics, Southern Utah University, Cedar City, UT 84720

Abstract

The malate synthase enzyme from Sulfolobus acidocaldarius has been reported to not require magnesium ions for activity, unlike all other examples of malate synthase which have been characterized. However, recent structural predictions of this enzyme using AI algorithms AlphaFold and RoseTTAFold indicate that the magnesium-binding residues found in all previously determined structures of malate synthases are also found in this enzyme, along with two additional key catalytic residues in a similar spatial constellation within an apparently conserved TIM barrel fold. To verify these structural predictions and to assess their accuracy, as well as confirm the divalent metal ion requirement of this enzyme, we plan to obtain an experimental structure using X-ray crystallography or Cryo-EM. Here we report cloning of the gene for the malate synthase from S. acidocaldarius into an overexpression vector using PCR and Gibson assembly, its transformation into E. coli, and overexpression of the protein.

†Corresponding authors: howard@suu.edu; elizabethpierce2@suu.edu *Undergraduate researchers and co-authors
Keywords: Protein Structure; Magnesium; Glyoxylate Cycle; Malate synthase; Thermophile; *Sulfolobus acidocaldarius*
Received: May 21, 2025 Accepted: May 27, 2025 Published: May 27, 2025

Introduction

Malate synthase is an enzyme integral to the glyoxylate cycle, an important metabolic hub for some microbes when living on two-carbon organic sources.[1] It allows the conversion of two acetyl groups to malate, a citric acid cycle intermediate.[2] The glyoxylate cycle also constitutes an essential drug target for pathogenic microbes such as *Candida albicans* and *Mycobacterium tuberculosis*.[3, 4]

Malate synthase is the second step in the conversion of isocitrate and acetyl-CoA to succinate and malate. Isocitrate lyase first splits isocitrate into succinate and glyoxylate. Malate synthase then utilizes this glyoxylate and acetyl-CoA to form malate and CoA. **(Figure 1)**

The proposed catalytic mechanism shown in Figure 1 emphasizes the importance of a magnesium ion in the catalytic cycle. It is essential for binding glyoxylate and polarizing it for attack by the enolate of acetyl-CoA. This Mg^{2+} is positioned in the active site by two acidic residues (Glu and Asp) to which it is tightly bound (not shown in Figure 1). The arginine and aspartate residues shown in Figure 1 are the other key groups that catalyze the reaction. These Arg and Asp residues, and the two Mg^{2+} binding residues, are conserved in all malate synthases which have been structurally determined previously using experimental methods.[6]

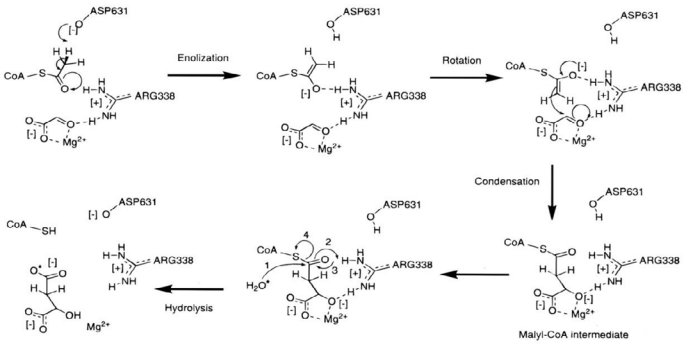

Figure 1. A currently proposed enzymatic mechanism for malate synthase with *E. coli* MSG numbering.[5]

The available experimental structures of the malate synthases have been structurally determined using X-ray crystallography, NMR, and cryo-EM. All of these belong to one of three classes: G(~730 residues in length), A(~530 residues), and H(~430 residues).[6] However, the malate synthase from *S. acidocaldarius* belongs to a fourth class that is significantly larger than these other classes(~830 residues), and has only been structurally predicted using artificial intelligence methods, AlphaFold[7] and RoseTTAFold.[8] The structures of the active site using both predictive methods are closely superimposable and have been described in detail.[9] These analyses concluded that the four key catalytic residues described above are structurally conserved, including the two Mg^{2+} binding residues, and therefore predicted that magnesium ions are essential for catalytic activity. This however appears to conflict with a report that this enzyme does not require magnesium for activity.[10] This paper also reported this enzyme formed a dimer in its active form. A dimer was therefore also modeled using ClusPro[11] with monomers predicted by RoseTTAFold.[9]

These structures of the *S. acidocaldarius* malate synthase are however only predictions. Our current project is to obtain an experimentally determined structure of this enzyme so the accuracy of these predictions can be assessed, and to verify that Mg^{2+} ions are indeed required for activity.

Experimental Methods

Growth of Sulfolobus acidocaldarius cultures

A lyophilized sample of the type strain (ATCC 33909/DSM 639) of *S. acidocaldarius*, originally isolated from Locomotive Spring in Yellowstone National Park, was obtained from the American Type Culture Collection.[12] Cells were resuspended and grown in the ATCC-recommended 1723 Revised Sulfolobus medium which contained 9.8 mM $(NH_4)_2SO_4$, 2.06 mM KH_2PO_4, 1.0 mM $MgSO_4 \cdot 7H_2O$, 0.48 mM $CaCl_2 \cdot 2H_2O$, 74 μM $FeCl_3 \cdot 6H_2O$, 9.14 μM $MnCl_2 \cdot 4H_2O$, 19 μM $Na_2B_4O_7$, 7.7 μM $ZnSO_4 \cdot 7H_2O$, 0.3 μM $CuCl_2 \cdot 2H_2O$, 0.12 μM $NaMoO_4 \cdot 2H_2O$, 0.15 μM $VOSO_4 \cdot 2H_2O$, 65 nM $CoSO_4$, 0.1% Tryptone, and 0.005% Yeast extract. The pH was adjusted to 3.0 with H_2SO_4, and filter-sterilized at 0.22 μm.

Cultures were grown in Fernbach flasks at 70 ºC, which was the maximum temperature for our shaking incubator, and 120 rpm. To prevent media from evaporating, a flask containing deionized water was placed in the shaker along with the culture flasks to humidify the incubator. Cells were collected for storage by spinning 20 min. at 5400 rpm and 4 ºC and stored at -80 ºC.

PCR amplification of the malate synthase gene

PCR primers were designed to amplify the malate synthase gene from genomic DNA from *S. acidocaldarius* cells. To obtain genomic DNA, a small portion of frozen *S. acidocaldarius* cell pellet was added to 50 µl of deionized water and boiled at 100 ºC for 15 minutes. This was added directly to the PCR reactions to provide template DNA. The gene was amplified from genomic DNA using the following primers (5' to 3'): forward-ACT TTA AGA AGG AGA TAT ACC ATG CCC TCC AAA TTA AAA ATT C; reverse- GTC GAC GGA GCT CGA ATT CGT TAT CAT TTA TAA TCA TCG TAT ATG TAG TC. The primer stocks were diluted to 20 µM and stored at -20 ºC. 20 µl PCR reactions were set up as follows (reactants added in order as listed): 1 µl nuclease-free H_2O, 4 µl 5x Phusion HF buffer, 2 µl of 2 mM dNTP mix, 5 µl of 2 µM forward primer, 5 µl of 2 µM reverse primer, 1 µl template DNA, 2 µl Phusion DNA polymerase HF (2 u/µl Phusion polymerase stock was first diluted 10-fold in nuclease-free water). For reactions of greater volume (50 µl or 100 µl), the reactants were scaled appropriately. A control reaction containing no template DNA was set up identically to the previously described reactions with nuclease-free H_2O used instead of template. The PCR reactions were then run on the thermocycler as follows: 98 °C for 1 minute followed by 35 rounds of 98 °C for 10 seconds, annealing temperatures (58.1 °C and 60.4 °C) for 15 seconds each, and 72 °C for 1 minute and 15 seconds, and finally 72 °C for 7 minutes before cooling to 4 °C until removed. PCR reactions were run in a 1% agarose gel at 90 volts for 1.5 hours to isolate the amplified DNA. GeneRuler 1kb Plus DNA ladder (ThermoFisher Scientific) was used to estimate sizes of the PCR products. The DNA band was then cut out of the gel and purified using the Monarch DNA Gel Extraction kit (NEB, cat: T1020S) following the provided protocol. The DNA concentration was measured at A_{260} using a Nanodrop spectrophotometer.

Plasmid construction using Gibson Assembly

Copies of the pET28b plasmid were made using competent DH10β *E. coli* cells. DH10β competent cells and pET28b plasmid stocks were thawed on ice. 50 µl of thawed cells were placed in a microfuge tube and 1 µl of plasmid was added to the cells. The mixture sat on ice for 30 minutes and was then heat shocked for 30 seconds in a water bath at 42 °C and placed back on ice to cool. Then 250 µl SOC media were added and the cells were incubated for 1 hour at 37 °C and 135 rpm. The cells were then plated on LB agar plates with 50 µg/ml kanamycin and placed in the incubator at 37 °C overnight. Individual colonies were then selected and grown in Terrific Broth (24 g/L yeast extract, 12 g/L tryptone, 4 mL/L glycerol, 100 mM KH_2PO_4 pH 5.0) containing 50 µg/ml kanamycin to an OD of about 3. The plasmid was then purified using the Monarch Plasmid DNA Miniprep kit (NEB, cat: T1010) and the DNA concentration was measured using a Nanodrop spectrophotometer.

The plasmid was then double-digested using NcoI and BamHI restriction enzymes. The following reaction was set up in a microfuge tube: 42 µl of purified pET28b plasmid (26 µl of 66.1 ng/µl and 16 µl of 97.1 ng/µl), 1 µl NcoI HF, 1 µl of BamHI HF, 5 µl cut-smart buffer and 1 µl nuclease-free H_2O. The contents were placed in a 37 °C water bath for 20 minutes and the reaction was stopped by adding 10 µl of 6X DNA loading dye. The digested plasmid was then run in a 0.8% agarose gel for 1.5 hours and the plasmid band was cut out of the gel and purified using the Monarch DNA Gel Extraction kit (NEB, cat: T1020S). GeneRuler 1kb Plus DNA ladder (ThermoFisher Scientific) was used to calibrate the gel. The DNA concentration was measured using a Nanodrop spectrophotometer.

The Gibson assembly reaction was set up by combining 0.35 µl of the purified MS gene PCR product DNA (174.1 ng/µl), 0.90 µl NcoI/BamHI digested pET28b plasmid (133.3 ng/µl), and 3.75 µl Gibson master mix. Gibson master mix was made by combining the following: 37 µl Taq ligase (40 u/µl), 100 µl 5x isothermal buffer, 2 µl T5 exonuclease (1 u/µl), 6.25 µl Phusion DNA polymerase (2 u/µl), and 229.75 µl nuclease-free H_2O. Two additional controls were set up with virgin pET28b, and NcoI/BamHI double digested pET28b. The virgin plasmid control reaction was set up by combining 1.25 µl of virgin pET28b plasmid (97.1 ng/µl) and 3.75 µl Gibson master mix. The double digest control was set up by combining 0.35 µl nuclease-free H_2O, 0.90 µl NcoI/BamHI double digested pET28b (133.3 ng/µl), and 3.75 µl Gibson master mix. The reactions were incubated at 50°C for 1 hour. Then 5 µl of each DNA reaction was added to 50 µl of DH10β competent cells and the mixture incubated on ice for 30 minutes. The cells were then heat shocked for 30 seconds in a 42°C water bath and placed back on ice to cool. 250 µl SOC media were then added, and the cells were placed in an incubator at 37°C and 135 rpm for 1 hour. The cells were then plated on LB agar plates containing 50 µg/ml kanamycin and placed in an incubator at 37°C overnight. The following day, six individual colonies were selected from the experimental plate and grown in Terrific broth to an OD of about 3. The plasmid from each growth was then purified using the Monarch Plasmid DNA Miniprep Kit (NEB, cat: T1010) and DNA concentration was measured using a Nanodrop spectrophotometer.

The purified plasmid samples were run on a 1% agarose gel at 90 V for 1.5 hours. GeneRuler 1kb Plus DNA ladder (ThermoFisher Scientific) was used to estimate sizes of the purified plasmids. Based on the increase in size, the plasmids from colonies 1, 2, 3, and 6 had successfully incorporated the gene. These bands were cut out of the gel and purified using the Monarch DNA Gel Extraction kit (NEB, cat: T1020S). DNA concentration was measured using a Nanodrop spectrophotometer. Each of these plasmids (named pSaciMS1, pSaciMS2, pSaciMS3, and pSaciMS6) was then verified through sequencing performed by GeneWiz.[13] The following internal sequencing primers were used (5' to 3'): 580- GAA GGC ACC AAA TCC TTA TTA CG; 1086- ATG GCT TAC CAG AAA TAT AAT GCG; 1630- ATA AGG AGT TAT GGG GAG GAA AG. T7 primers provided by GeneWiz were used to cover the ends of the gene.

Protein overexpression

The plasmid pSaciMS6 was chosen to be transformed into protein expression competent cells. Rosetta™(DE3) competent cells (Novagen) were thawed, and 0.5 µl pSaciMS6 (121.6 ng/µl)

was added to 50 µl of Rosetta cells. The cells sat on ice for 30 minutes, were heat shocked for 30 seconds in a 42°C water bath, then placed back on ice to cool. 250 µl SOC media was added to the cells, and they were incubated at 37°C and 135 rpm for 1 hour. The cells were then plated on LB agar plates containing 50 µg/ml kanamycin and 25 µg/ml chloramphenicol and placed in an incubator at 37°C overnight. The following day, a single colony was selected and grown in LB media containing 50 µg/ml kanamycin and 25 µg/ml chloramphenicol. Cells were frozen in 25% glycerol at -80°C.

From a frozen stock of transformed Rosetta cells, four tubes with 5 ml of LB media with 50 µg/ml kanamycin and 25 µg/ml chloramphenicol were inoculated and grown overnight at room temperature with no shaking. These were then added to four Fernbach flasks each containing 1 L of LB media with 50 µg/ml kanamycin and 25 µg/ml chloramphenicol and the cultures were placed in a shaking incubator at 37°C and 120 rpm. When cells had grown to an OD of 0.55, a sample was removed as a pre-induction reference to compare post-induction protein expression, and 1 ml of 1M IPTG was added to each flask. The sample removed for protein expression monitoring was spun at 13,500 rpm for 1 minute. The pellet was resuspended in 5X SDS loading buffer and boiled at 90°C for 15 minutes. A sample was taken every hour for 4 hours to analyze the induction of protein expression. The four liter-cultures were then spun down at 5400 rpm and 4°C for 10 minutes and the cell pellet was frozen at -80°C. The samples to monitor overexpression were then run in an SDS PAGE gel (12% for resolving and 5% for stacking) for 15 minutes at 120V followed by 45 minutes at 200V. The gel was then stained with Coomassie Blue and imaged on a BioRad ChemiDoc XRS+ gel imager.

Results and Discussion

Growth of Sulfolobus acidocaldarius

Attempts to grow cultures of *S. acidocaldarius* to high densities for effective production of protein were unsuccessful as this environmental strain plateaued at relatively low cell densities when grown at the limiting temperature of the shaker-incubator in our lab. Therefore, we decided to clone the gene for overexpression using the cell pellets of *S. acidocaldarius* which we used as a source of genomic DNA.

Polymerase Chain Amplification of Gene

Using PCR,[14] the gene was amplified from *S. acidocaldarius* genomic DNA. **(Figure 2)** The set of designed primers effectively amplified a single product of the expected size with virtually no other products formed. The PCR amplification product is expected to be 2519 bp in length. A set of nested primers designed as a back-up plan also selectively amplified the local gene region from chromosomal DNA.

Gibson Assembly

The PCR product was cut out of the gel using a scalpel, and gel-purified. A sample of pET28b plasmid was doubly digested using NcoI/BamHI in order to produce the planned ends to allow annealing of the PCR product in the proper orientation. Gibson Assembly[15] was used to build the expression plasmid. Gibson master mix was added and the assembly was carried out as described in the methods section. Six colonies were selected from the plated Gibson reaction and grown in overnight cultures. Samples of each product from plasmid minipreps were run on an agarose gel. **(Figure 3)** It is clear from the gel shift compared to the original virgin pET-28b in lane 5 that colony #4 (lane 6) and colony #5 (lane 7) did not incorporate the malate synthase gene which should have added about 2.4 kb to the plasmid size. The virgin pET-28b is 5369 bp, and pSaciMS should have had 101 bp cut out with the double digest and 2478 bp added (with an additonal stop codon) for an expected size of 7746 bp for the successful, planned expression construct. This larger size is clearly seen in the gel shifts in lanes 2, 3, 4, and 8. The plasmid in lane 8 (pSaciMS6) was chosen for overexpression experiments.

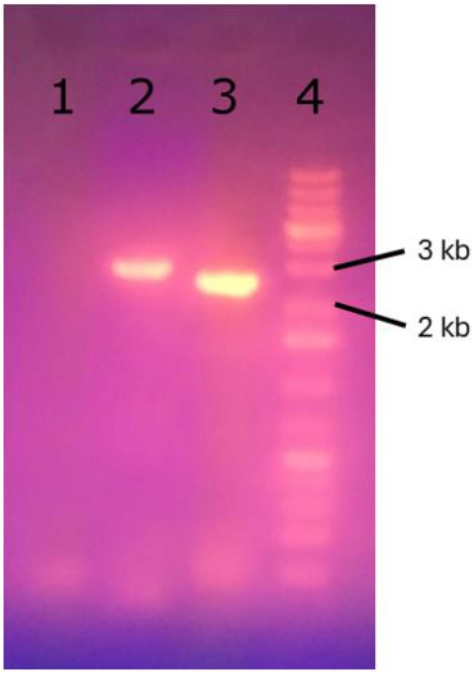

Figure 2. Agarose gel of PCR amplification of *S. acidocaldarius* gene from genomic DNA. Lane 1: Control with no template DNA, Lane 2: Nested set of primers, Lane 3: Primers designed to link gene into overexpression plasmid, Lane 4: GeneRuler 1 kb Plus DNA ladder

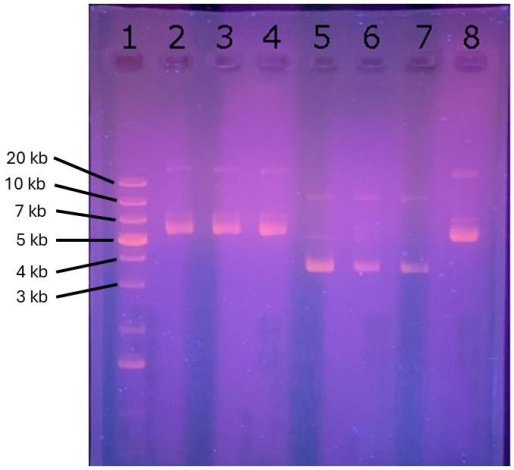

Figure 3. Agarose gel of the products of Gibson Assembly. Lane 1: GeneRuler 1 kb Plus ladder, Lane 2: pSaciMS1, Lane 3: pSaciMS2, Lane 4: pSaciMS3, Lane 5: Virgin pET28b, Lane 6: plasmid from colony #4, Lane 7: plasmid from colony #5, Lane 8: pSaciMS6

Overexpression of the malate synthase enzyme

Due to the large number of rare codons in the *S. acidocaldarius* gene compared to the codon preferences of *E. coli*, we used an *E. coli* strain containing the pRARE plasmid,[16] under chloramphenicol selection, which expresses the rare tRNAs (for codons: AUA, AGG, AGA, CUA, CCC, and GGA) that would be needed in large numbers for efficient overexpression. The *S. acidocaldarius* malate synthase gene contains 19 AGG and 23 AGA codons for Arginine, 12 CCC codons for Proline, and 26 GGA codons for Glycine. Therefore the pSaciMS6 plasmid (lane 8, Figure 3) was transformed into a Rosetta™ (DE3) cell line which contains pRARE. The cells were grown under double selection of kanamycin and chloramphenicol to maintain the pSaciMS6 and pRARE plasmids respectively. Four 1-liter cultures were grown to early log phase (A_{600}= 0.55) at 37 °C and 120 rpm and then induced with 1 mM IPTG. Expression of the protein was analyzed by SDS PAGE and Coomassie staining **(Figure 4)** with samples taken immediately before, and then every hour after induction for four hours total.

A single induction band is clearly seen to appear after one hour following induction with IPTG. And this band continues to intensify over time building to high levels by four hours postinduction. It is interesting that this band is running faster than is expected for its size. It looks to be running slightly ahead of the 85 kDa band in the ladder, but each complete monomer should be 94.65 kDa. Aberrant mobility of proteins through an SDS PAGE is not entirely unexpected, and given the expected error of ±10% in molecular weights using this technique,[17] its migration is consistent with the expected protein size. Importantly, before selecting pSaciMS6 for overexpression experiments the plasmid was sequenced to confirm that there were no mutations from the wild-type sequence, and therefore there are no premature stop codons.

Acknowledgements

The authors thank the L.S. and Aline W. Skaggs Research Endowment, and the Walter Maxwell Gibson Endowment at SUU for providing financial support for this work.

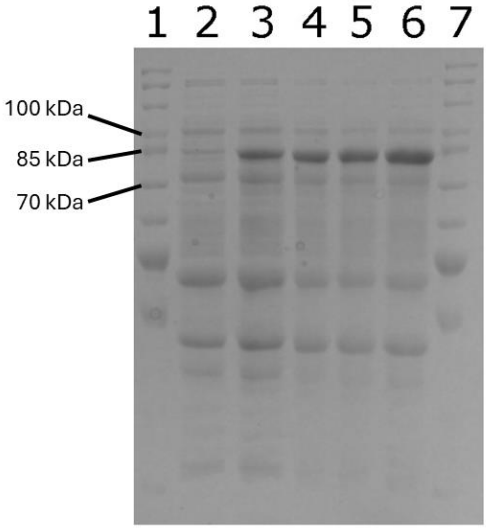

Figure 4. SDS-PAGE of *S. acidocaldarius* malate synthase overexpression. Lane 1: PageRuler Protein Ladder (Sizes shown are 200, 150, 120, 100, 85, 70, 60, 50, and 40 kDa) Lane 2: Preinduction, Lane 3: 1 hr. postinduction, Lane 4: 2 hrs. postinduction, Lane 5: 3 hrs. postinduction, Lane 6: 4 hrs. postinduction, Lane 7: same as lane 1.

References

1. Kornberg, H. L. *Biochem. J.,* **1966**, 99(1), 1-11.
2. Kornberg, H. L.; Krebs, H. A. *Nature,* **1957**, 179, 988-991.
3. McKinney, J. D.; Höner zu Bentrup, K.; Muñoz-Elías, E. J.; Miczak, A.; Chen, B.; Chan, W. T.; Swenson, D.; Sacchettini, J. C.; Jacobs, W. R., Jr.; Russell, D. G. *Nature,* **2000**, 406(6797), 735–738.
4. Lorenz, M. C.; Fink, G. R. *Nature,* **2001**, 412, 83-86.
5. Howard, B. R.; Endrizzi, J. A.; Remington, S. J. *Biochemistry,* **2000**, 39(11), 3156–3168.
6. Bracken, C. D.; Neighbor, A. M.; Lamlenn, K. K.; Thomas, G. C.; Schubert, H. L.; Whitby, F. G.; Howard, B. R. *BMC Struct. Biol.,* **2011**, 11(23), 1-19.
7. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. *Nature,* **2021**, 596, 583-589.
8. Baek, M.; DiMaio, F.; Anishcenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch L. N.; Schaeffer, R. D.; Millán. C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. *Science,* **2021**, 373:6557, 871-876.
9. Nielsen, S.; Orton, J.; Howard, B. R. *AJUR,* **2024**, 20(4), 29–37.
10. Uhrigshardt, H.; Walden, M.; John, H.; Petersen, A.; Anemüller, S. *FEBS. Lett.,* **2002**, 513(2-3), 223–229.
11. Yueh C.; Hall D. R.; Xia B.; Padhorny D.; Kozakov D.; Vajda S. *J. Mol. Biol.,* **2017**, 429(3), 372-381.
12. American Type Culture Collection (ATCC). P.O. Box 1549, Manassas, VA 20108, USA. https://www.atcc.org/.
13. GeneWiz, 115 Corporate Blvd. South Plainfield, NJ 07080, USA. https://www.genewiz.com/.
14. Saiki, R.K.; Gelfand, D.H.; Stoffel, S.; Scharf, S.J.; Higuchi, R.; Horn, G.T.; Mullis, K.B.; Erlich, H.A. (1988) *Science,* **1988**, 239(4839):487-91.
15. Gibson, D.G.; Young, L.; Chuang, R.Y.; Venter, J.C.; Hutchison, C.A. 3rd; Smith, H.O. *Nat. Methods,* **2009**, 6(5):343-5.
16. Novy, R.; Drott, D.; Yaeger, K.; Mierendorf, R. *inNovations,* **2001**, 12, 1-3.
17. Hames, B.D. *Gel Electrophoresis of Proteins: A Practical Approach (Practical Approach Series)* 3rd Edition. Oxford University Press, Oxford, NY **1998** pp. 29-33.