

Identification of key functional sites for ATPase activity in the OBC protein complex using artificial intelligence approaches

Lei Hua(花蕾), Yuhang He(贺宇航), Peng Zhao(赵澎), Jiayao Liu(刘嘉瑶), and Kun Shang(尚坤)
Citation: Chin. Phys. B, 2026, 35 (**4**): 048702. DOI: 10.1088/1674-1056/ae0891
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Force-dependent unfolding dynamics of spectrin R16: Resolving experimental contradiction and unveiling model consistency

Wanxing Zhang(张万星), Zhuwei Zhang(张珠伟), Zhenyong Xue(薛振勇), Yuhang Zhang(张宇航), Shimin Le(乐世敏), and Hu Chen(陈虎)
Chin. Phys. B, 2025, 34 (**8**): 088708. DOI: 10.1088/1674-1056/adbdbf

Effects of palmitoylation on membrane protein partitioning into lipid domains in model cellular membranes

Shishi Wu(吴施施) and Qing Liang(梁清)
Chin. Phys. B, 2025, 34 (**5**): 058701. DOI: 10.1088/1674-1056/adbee9

Effect of chaperone-client interaction strength on Hsp70-mediated protein folding

Lujun Zou(邹禄军), Jiajun Lu(陆伽俊), and Xiulian Xu(徐秀莲)
Chin. Phys. B, 2023, 32 (**11**): 118701. DOI: 10.1088/1674-1056/acea6f

Single-molecular methodologies for the physical biology of protein machines

Shuang Wang(王爽), Ying Lu(陆颖), and Ming Li(李明)
Chin. Phys. B, 2022, 31 (**12**): 128702. DOI: 10.1088/1674-1056/ac98a2

Evaluation on performance of MM/PBSA in nucleic acid-protein systems

Yuan-Qiang Chen(陈远强), Yan-Jing Sheng(盛艳静), Hong-Ming Ding(丁泓铭), and Yu-Qiang Ma(马余强)
Chin. Phys. B, 2022, 31 (**4**): 048701. DOI: 10.1088/1674-1056/ac3a5c

---

Identification of key functional sites for ATPase activity in the OBC protein complex using artificial intelligence approaches

Lei Hua(花蕾)[1,4,†], Yuhang He(贺宇航)[1,2,3,†], Peng Zhao(赵澎)[1,†], Jiayao Liu(刘嘉瑶)[1], and Kun Shang(尚坤)[1,‡]

[1] *Yan'an Medical College of Yan'an University, Yan'an* 716099, *China*
[2] *Beijing National Laboratory for Condensed Matter Physics and Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing* 100190, *China*
[3] *Songshan Lake Materials Laboratory, Dongguan* 523830, *China*
[4] *NHC Key Laboratory of Systems Biology of Pathogens, National Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing* 102629, *China*

(Received 25 April 2025; revised manuscript received 12 September 2025; accepted manuscript online 18 September 2025)

Microorganisms evolve diverse immune defense systems to protect against phage and viral invasion. The bacteriophage exclusion (BREX) system represents a common bacterial genomic defense mechanism, and preliminary studies indicate that the interaction between BrxC and OrbA proteins (forming the OBC complex) can alter the aggregation pattern of defense system components and modulate immune function. Notably, the adenosine triphosphatase (ATPase) domain of this system plays a pivotal role. This study employs optimized network screening (ONS), an artificial intelligence-based integrated analytical approach combining machine learning and deep learning, to identify critical enzymatic active sites in target macromolecular complexes and to provide crucial insights for potential drug target discovery. Through comprehensive ONS analysis of multiple sites, we identified lysine (K73) as a key residue governing the ATPase activity of this protein complex. Following plasmid construction and expression, we successfully purified the OrbA–BrxC (OBC) complex, along with newly engineered OBCK (K73A mutant of the OBC complex) and the negative control OBCE (E255A mutant of the OBC complex). ATPase activity assays demonstrated complete loss of enzymatic function in the OBCK mutant, confirming K73 as the essential catalytic site for BREX system ATPase activity. Comparative analysis revealed minimal differences between OBC and OBCE complexes, but significant divergence from OBCK. The single K73 mutation alone was sufficient to abolish critical enzymatic activity and disrupt BREX system function, achieving efficient complex engineering without requiring multiple mutations. Building on this artificial intelligence methodology, we further analyzed potential functional sites and antigenic epitopes in OBC and OBCK complexes, thereby establishing a research foundation for developing anti-drug-resistant pathogen therapeutics.

Keywords: artificial intelligence methods, BREX defense system, co-expression and purification, ATPase activity, construction of protein complexes

PACS: 87.14.E–, 87.15.hp, 87.15.km, 87.50.sg

DOI: 10.1088/1674-1056/ae0891 **CSTR:** 32038.14.CPB.ae0891

1. Introduction

To defend against the invasion of bacteriophages and viruses, bacteria have evolved a variety of immune defense systems, including the clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated (Cas) system,[1–3] restriction–modification (R–M) systems,[4] and cyclic oligonucleotide-based anti-phage signaling systems (CBASS).[5] Goldfarb *et al.*[6] identified a novel phage exclusion mechanism in Bacillus cereus, termed the bacteriophage exclusion (BREX) system. The BREX system is relatively common, found in approximately 7% of prokaryotic genomes.[7] Similar to R–M systems, BREX can epigenetically modify the host chromosome, a feature thought to distinguish self-deoxyribonucleic acid (DNA) from foreign, unmodified DNA.[6,8] To date, six types of BREX systems have been described,[6] with type I being the most widespread. It typically consists of six open reading frames: brxA, brxB, brxC, brxX, brxZ, and brxL. brxA has been shown to possess structural features and DNA-binding activity;[9] BrxB lacks clearly defined structural domains; BrxC contains an adenosine triphosphatase (ATPase) domain identified through bioinformatic analysis;[6] BrxL shows homology to MCM replicative helicases found in archaea and eukaryotes.[10] BrxX functions as a DNA methyltransferase,[8] while BrxZ belongs to the alkaline phosphatase superfamily.[6] Type I BREX systems inhibit phage genome replication, though no degradation of phage DNA has yet been observed,[6,8] and the precise mechanism by which BREX blocks phage replication without cleaving phage DNA remains unclear. BREX systems play a critical role in maintaining bacterial physiological integrity. In Escherichia coli, BREX provides protection against exogenous double-stranded DNA contamination,[8] and in *Salmonella*, it

[†]These authors contributed equally to this work.
[‡]Corresponding author. E-mail: puospumu@yeah.net

© 2026 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

safeguards pathogenic strains via DNA modification.[11] With the growing emergence of antimicrobial resistance among pathogens such as *Salmonella*, *E. coli*, and *Vibrio cholerae*, these bacteria have become increasingly difficult to control, posing serious public health threats globally. Therefore, identifying universal targets within bacterial defense systems, such as BREX, holds great promise for combating drug-resistant bacterial infections.

Although the BREX defense system is common among bacteria, only three proteins have been identified to interfere with its activity: overcome classical restriction (Ocr),[12] S-adenosylmethionine (SAM) lyase,[13] and Ocr-related BREX protein A (OrbA).[14] The Ocr protein, encoded by bacteriophage T7, is the first phage-encoded protein discovered to eliminate BREX-mediated defense.[12] It is well known that Ocr inhibits R–M systems by acting as a DNA mimic.[12,15,16] Ocr can also directly bind to the DNA methyltransferase BrxX, thereby restoring phage propagation.[12] During T7 phage infection, Ocr specifically suppresses the BREX system, leading to a significant reduction in host genome methylation. T7 phages lacking Ocr or carrying mutations at the Ocr dimer interface have lost resistance to type I BREX systems,[15] suggesting that Ocr specifically inhibits the activity of the core component BrxX methyltransferase in the BREX system. Drobiazko *et al.*[16] found that the inhibitor Ocr binds with high affinity to BrxX, forming a stable tetrameric complex consisting of two BrxX monomers and an Ocr dimer. As a DNA mimic, Ocr sequesters the BrxX monomers, locking them in an inactive open conformation to prevent DNA binding, thereby successfully inhibiting the activity of the BREX system. The Escherichia coli T3 phage inhibits the BREX system by degrading SAM through SAM lyase, and also suppresses SAM synthesis by binding to the enzyme MetK.[13] SAM lyase has also been shown to protect against R–M systems,[17] suggesting that both R–M and BREX systems use similar defense mechanisms to protect themselves from interference. OrbA can partially protect the lytic phage ICP1 from BREX-mediated restriction in its host, the pathogenic *Vibrio cholerae*.[14] Previous studies have shown that OrbA can form complexes with BrxC in the BREX system to create the OrbA + BrxC (OBC) complex, which reduces ATPase activity and interferes with the aggregation pattern of proteins in the BREX system, thereby weakening its immune function.[18]

Currently, there are no reports of successfully constructing and purifying the key BREX-associated protein complex OBC. The formation of the OBC protein complex can inhibit the functionality of the BREX defense system, with its ATPase domain playing a crucial role in the defense mechanism. This complex reduces ATPase activity and alters the aggregation patterns of monomers within the system, thereby suppressing the BREX defense function.[18] Similar to other protein complexes in the BREX defense system, such as the complex formed by Ocr and BrxX, the formation of these complexes is primarily driven by electrostatic interactions between the positively charged regions of one monomer and the groove of another. BREX site recognition requires flipping of the target adenine, which transitions the complex from an open to a closed conformation, locking the monomers in an open state.[15] As a homologous system, OBC may operate via a similar mechanism, altering the aggregation pattern of monomers. Additionally, in the Gabija system, which also contains an ATPase domain, ATP binding to the ATPase domain of the protein complex formed by GajA and GajB stabilizes the inserted region within the domain, keeping the other domains in a closed conformation.[19] The OBC protein complex, similar to the Gabija complex, contains an ATPase domain that likely exerts analogous functional effects by modulating the domain through stabilization of insertion regions. If the ATPase domain of the BREX defense system is disrupted, the internal conformation of its active protein could be affected, potentially leading to suppression of the BREX system's function. Given the crucial role of the ATPase domain in the BREX defense system, precise identification of functional sites associated with this domain within the protein complex could provide strategic direction for modulating BREX immune function.

Systematic identification of functionally critical sites within the OBC protein complex presents a significant research challenge due to the abundance of potential amino acid targets. Notably, no prior studies have comprehensively analyzed key sites in the BREX defense system-associated OBC complex. In this study, we employ artificial intelligence-based approaches to, for the first time, map essential functional targets in the OBC protein complex. Currently, artificial intelligence (AI) techniques are advancing rapidly, with main methods including machine learning, neural networks, and deep learning.[20] This study adopts an AI approach that integrates machine learning and deep learning techniques to identify key target sites in the OBC protein complex. By utilizing data-driven models that learn from data and combining them with the large AI model AlphaFold3 for comparison and analysis, this study constructs a comprehensive analytical framework, optimized network screening (ONS), to achieve multi-dimensional analysis of protein features and to locate potential important sites. Specifically, ProtParam and ProtScale were employed to analyze biophysical properties and hydrophobicity patterns; transmembrane hidden Markov model (TMHMM) was utilized for transmembrane domain prediction; SignalP was applied for signal peptide identification; NetPhos was implemented for phosphorylation site prediction; SPOMA was integrated for secondary structure determination; and AlphaFold3 was used for tertiary struc-

ture modeling and comparative structural analysis (Fig. 1). These machine learning approaches, trained on large-scale datasets, can extract feature patterns from protein sequence data and enable high-precision analysis of structural and functional characteristics of the OBC protein complex. The three-dimensional structure model is systematically evaluated using various structural validation methods, such as error recognition analysis tool (ERRAT), Verify 3D, and Ramachandran plots, and the model with the highest quality score is selected. Furthermore, the ONS method is applied for comparative analysis of the numerous sites within the complex, enabling identification of key catalytic target sites in BREX-related macromolecules. This integrated approach offers a significant advantage by providing comprehensive and precise site selection through the synergy of multiple algorithms. By leveraging the complementary strengths of each method, it minimizes potential prediction errors associated with individual tools, thereby improving our understanding of complex protein functions.

Based on the previous analysis of the target protein complex using the ONS method, we first used molecular cloning techniques to construct the plasmid for the OBC protein complex and carried out its expression and purification. Following this, we constructed the OBCK and OBCE protein complexes based on the key target sites identified by the ONS method, and expressed and purified both complexes. Subsequently, we measured their ATPase activity. We found that the newly constructed OBCK protein complex completely lost its enzymatic activity, leading to disruption of the BREX system's defense function. We successfully used the ONS method, an artificial intelligence-based technique, to efficiently and accurately locate key target sites in the OBC protein complex, and validated a high-efficiency strategy that can abolish defense system function with a single mutation. Furthermore, through the ONS method, we analyzed potential drug target sites and antigen epitopes of this defense system protein complex, which will play a crucial role in the future development of drugs targeting drug-resistant pathogens.

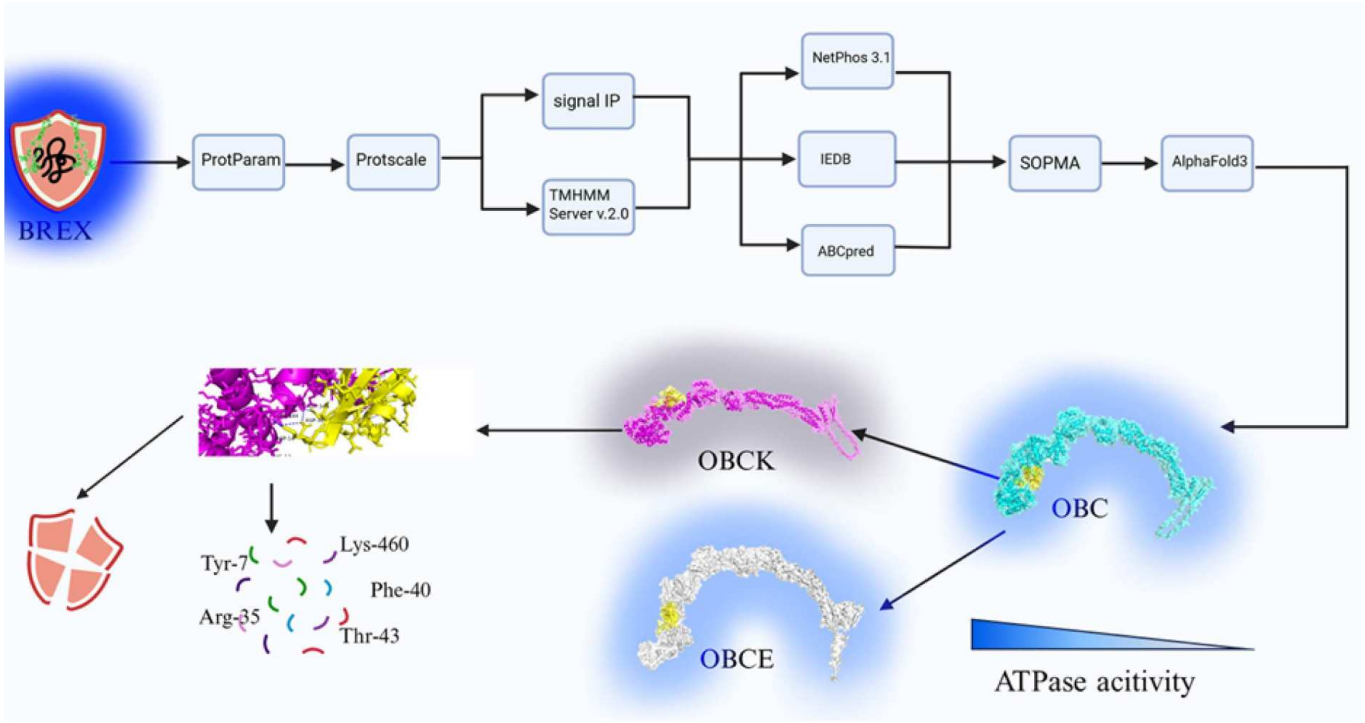


Fig. 1. Identification of key ATPase activity sites in the OBC protein complex based on artificial intelligence methods.

2. Results

2.1. Discovery of key ATPase activity sites in the OBC protein complex

The molecular formula of the OBC protein complex is $C_{6493}H_{10210}N_{1682}O_{1998}S_{24}$, composed of 20407 atoms, including five elements: C, H, O, N, and S. The relative molecular mass is 144574, with a theoretical isoelectric point of 5.40. The total number of positively charged amino acid residues (Arg + Lys) in the protein complex is 161, while the total number of negatively charged amino acid residues (Asp + Glu) is 189. When the N-terminal residue of the protein complex is methionine, the half-life of the complex in mammalian reticulocytes in vitro is 30 hours, in yeast cells is over 20 hours, and in Escherichia coli cells is over 10 hours. The instability index is less than 40 for stable proteins, and since the instability index of this protein complex is 35.53, it is considered a stable protein. The aliphatic index is 87.56. The hydrophobicity and hydrophilicity distribution of amino acids in the OBC protein complex is shown in Fig. 2(a). The polypeptide chain of the

complex has the highest hydrophobic value (2.556) at threonine at position 816 and the lowest hydrophilic value (−3.111) at aspartic acid at position 666. Overall, the complex is hydrophilic. The subcellular localization of the protein complex is the nucleus, indicating its presence within the cell nucleus and its involvement in cellular information storage and transcription. Phosphorylation sites refer to specific amino acid residues that can be phosphorylated by kinases, a modification that plays a central role in cellular signal transduction and metabolic regulation. The OBC protein complex contains 135 phosphorylation sites (Fig. 2(b)).

In the secondary structure of the OBC protein complex, α-helices (Hh) account for 841 residues (66.06%), random coils (Cc) for 351 residues (27.57%), and extended strands (Ee) for 81 residues (6.36%) (Fig. 2(c)). We next utilized an artificial intelligence-based model to construct and analyze the structural model of the OBC protein complex (Fig. 2(d)). In this study, we employed several tools to validate the predicted protein structure and ensure its high quality, including ERRAT, VERIFY3D, and the Ramachandran plot. The overall quality factor of the protein structure, calculated by ERRAT, was 93.078, indicating good stereochemical quality (Fig. S1). VERIFY3D analysis showed that 80.73% of the backbone residues had a 3D–1D average score greater than 0.1, meeting the criterion of at least 80% passing. This suggests that the 3D structure of the protein is consistent with its primary sequence, supporting the quality of the model (Fig. 2(e)). Additionally, Ramachandran plot analysis revealed that 94.11% of the residues were in the most favored regions, 5.7% in additionally allowed regions, 0.2% in generously allowed regions, and none in disallowed regions (Fig. 2(f)). These results, obtained from different validation tools, consistently demonstrate that the protein structure is of high quality and confirm the reliability of the model.

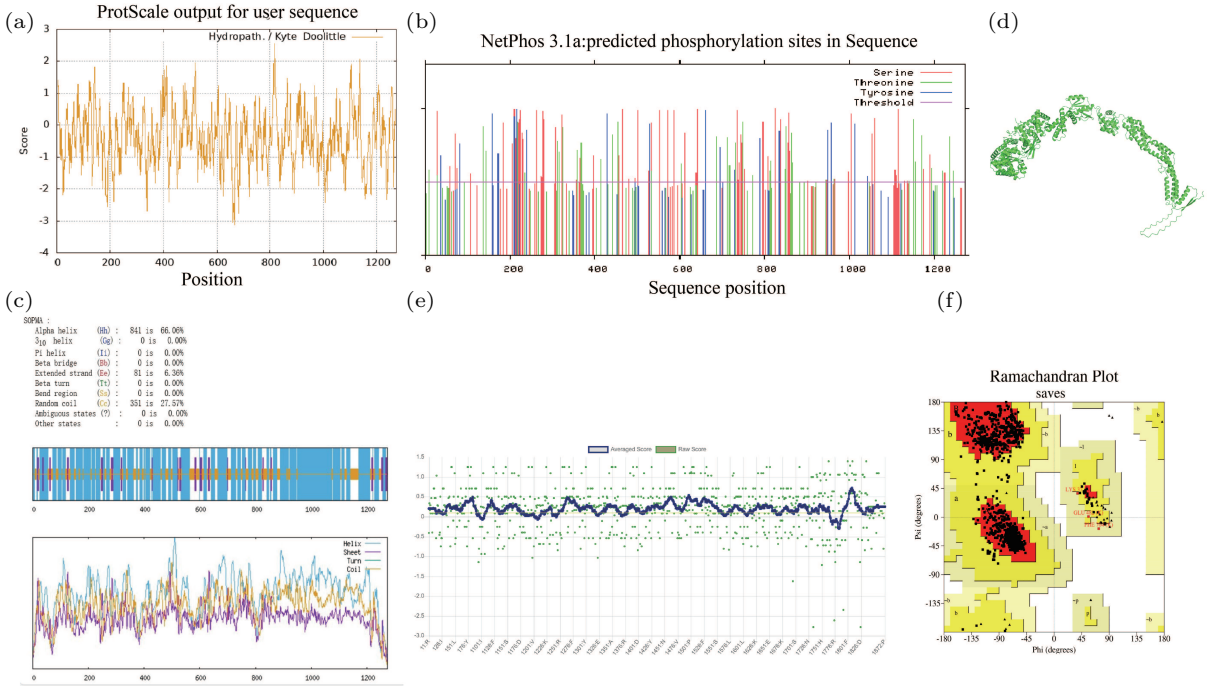


Fig. 2. Identification of key functional sites in the OBC protein complex. (a) Prediction of hydrophilicity and hydrophobicity of the OBC protein complex. (b) Prediction of phosphorylation sites in the OBC protein complex. (c) Prediction of the secondary structure of the OBC protein complex. (d) Prediction of the tertiary structure of the OBC protein complex. (e) Verify 3D results for the OBC protein complex. (f) Ramachandran plot of the OBC protein complex. Glycine residues are represented by triangles, while other amino acid residues are represented by squares. The plot indicates the following regions: most favored regions (A, B, L, red); additional allowed regions (a, b, l, p, yellow); generously allowed regions (∼ a, ∼ b, ∼ l, ∼ p, light yellow); disallowed regions (white).

2.2. Construction, expression, and purification of the OBC protein complex

Next, we constructed the OBC protein complex. The obtained sequence was integrated into the target vector through homologous recombination to construct the plasmid. The synthetic plasmid was analyzed by agarose gel electrophoresis to determine its length. The electrophoresis results showed that the length of the synthetic plasmid was the same as that of the original plasmid, indicating successful construction of the OBC protein complex plasmid (Fig. 3(a)). The frozen OBC protein complex bacterial pellet was resuspended in lysis buffer containing 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 10 mM imidazole, and 1 mM PMSF. After 20 minutes of sonication and high-speed centrifugation, the supernatant was collected and subjected to Ni–NTA affinity purification. The protein was washed with lysis buffer to remove contaminants and eluted with a buffer containing 20 mM Tris–HCl (pH 8.0), 50 mM NaCl, and 300 mM imidazole. The eluate was collected, and 20 μL of the eluted solution was analyzed by SDS–PAGE. The results showed a protein band at the ex-

pected position, confirming the preliminary purification of the OBC protein complex (Fig. 3(b)).

Ion-exchange chromatography is a technique that separates proteins based on differences in surface charge. It utilizes electrostatic interactions between charged molecules in the mobile phase buffer and ion-exchange media in the stationary phase to achieve separation during adsorption and desorption processes. We co-expressed the two proteins of the OBC complex in competent cells, followed by IPTG induction and Ni–NTA purification, including washing and elution steps. The eluted proteins were then subjected to ion-exchange chromatography. SDS–PAGE analysis revealed the presence of protein bands at the expected positions, indicating the successful purification of the OBC protein complex (Fig. 3(c)).

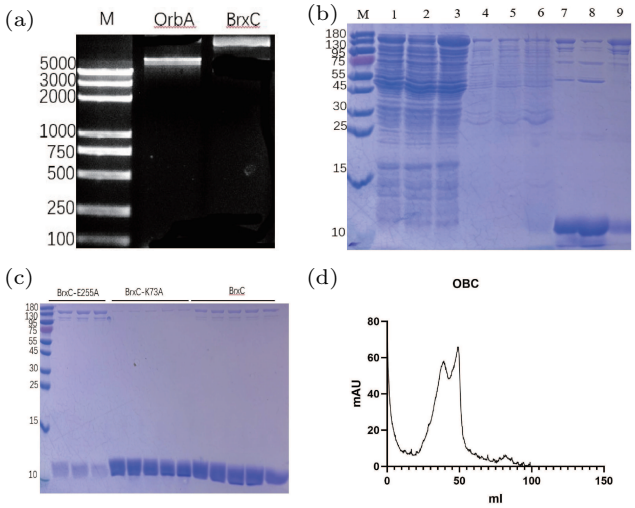


Fig. 3. Expression, construction, and purification of the OBC protein complex. (a) Identification of the OBC protein complex plasmid. (b) Purification of the two proteins of the OBC protein complex. M: marker; 1: supernatant of bacterial lysate expressing the OBC protein complex; 2: supernatant of bacterial lysate expressing the OBCK protein complex; 3: supernatant of bacterial lysate expressing the OBCE protein complex; 4: pellet after high-speed centrifugation of OBC protein complex bacteria; 5: pellet after high-speed centrifugation of OBCK protein complex bacteria; 6: pellet after high-speed centrifugation of OBCE protein complex bacteria; 7: elution of OBC protein complex from Ni–NTA; 8: elution of OBCK protein complex from Ni–NTA; 9: elution of OBCE protein complex from Ni–NTA. (c) Ion-exchange chromatography results of the OBC protein complex. (d) Ion-exchange chromatogram of the OBC protein complex.

2.3. Novel construction of the OBC protein complex and measurement of ATPase activity

To identify key functional residues in OBC, we first analyzed its hydrophilicity and hydrophobicity using ProtScale, which revealed 869 hydrophilic sites (Fig. 2(a)). Next, we predicted potential phosphorylation sites using NetPhos and excluded already phosphorylated residues, leaving 710 candidate sites (Fig. 2(b)). In addition, analyses with TMHMM and SignalP indicated that OBC contains neither a signal peptide nor transmembrane regions (Figs. S2 and S3). We then performed a cross-species sequence alignment to identify conserved residues across different bacterial species (Fig. S4). The conserved sites were overlapped with the remaining candidate sites, ultimately yielding 35 highly conserved residues. These sites were integrated with the AlphaFold3-predicted 3D structure of OBC, with particular focus on residues located at the beginning and end of secondary structures, as these positions are often crucial for maintaining secondary structure stability. Through this integrated approach, we identified five candidate residues — 62G, 73K, 186W, 314R, and 321L. Based on the proximity of these residues to the ATP-binding site, K73 was identified as the nearest residue, thereby supporting its role as a potential key functional site. This is related to the Walker A motif described in the literature.[20,21] The key component of the OBC protein complex contains a conserved Walker A motif (GXXXXGKT), which is commonly found in proteins with ATP hydrolytic activity. The Walker A motif primarily functions in ATP binding. To investigate the role of this motif, we mutated the critical lysine (K) residue in the Walker A sequence to alanine (A), thereby disrupting ATP binding capability. Additionally, we targeted the Walker B motif (hhhhDE) — which is not directly involved in ATP binding but is essential for ATP hydrolysis — by mutating its key glutamate (E) residue to alanine (A), thereby impairing ATP hydrolysis. These mutations were introduced into BrxC, generating the BrxC–K73A and BrxC–E255A variants. We amplified the BrxC–K73A and BrxC–E255A sequences using PCR, and the resulting products were analyzed by 1% agarose gel electrophoresis to confirm their lengths. The electrophoresis results showed that both mutant fragments matched the expected lengths of the original sequences, indicating that the mutant constructs were successfully generated (Fig. 4(a)).

We subsequently co-expressed the successfully constructed BrxC mutants with OrbA to generate the OBCK and OBCE protein complexes. Bacterial pellets were resuspended in lysis buffer (20 mM Tris–HCl, pH 8.0; 150 mM NaCl; 10 mM imidazole; 1 mM PMSF), followed by ultrasonication for 20 minutes. After high-speed centrifugation, the supernatant was collected and subjected to Ni–NTA affinity purification. Impurities were removed using wash buffer (same as lysis buffer), and target proteins were eluted with elution buffer (20 mM Tris–HCl, pH 8.0; 50 mM NaCl; 300 mM imidazole). Eluted samples (20 μL) were analyzed by SDS–PAGE. As shown in Fig. 3(b), both OBCK and OBCE protein complexes were successfully purified.

The ATPase domain is a core module for cellular energy metabolism and mechanical function, characterized by a highly conserved sequence and structure. We performed ATPase activity assays on the three protein complexes — OBC, OBCK, and OBCE. The results showed that the OBC protein complex exhibited a low level of ATPase activity, while the OBCK protein complex completely lost ATPase activity. Interestingly, the OBCE protein complex maintained its ATPase

activity (Fig. 4(b)). These results indicate that lysine in the OBC protein complex is the key residue for ATPase activity, and that mutation of this residue completely abolishes the ATPase activity of the OBCK complex.

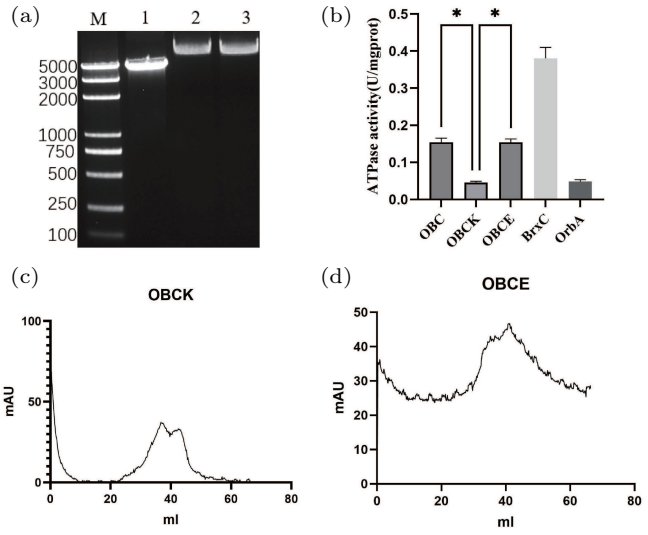


Fig. 4. Modification of OBC and ATPase activity. (a) Plasmid identification of OBCK and OBCE protein complexes. M: marker; 1: pET-Duet vector; 2: BrxC–K73A; 3: BrxC–E255A. (b) ATPase activity of OBC, OBCK, and OBCE protein complexes. Statistical differences between groups were analyzed using a t-test; $p < 0.05$ $(*)$, $p < 0.01$ $(**)$, $p < 0.005$ $(***)$. (c) Ion-exchange chromatography profile of the OBCK protein complex. (d) Ion-exchange chromatography profile of the OBCE protein complex.

To further validate the specificity of K73, we constructed three additional mutants (R35A, R38A, and T915A) and purified their proteins. None of these mutations abolished ATPase activity to the extent observed in OBCK (Fig. S5A). Intriguingly, while the wild-type OBC complex inherently suppresses ATPase activity (likely due to BrxC–OrbA interactions), the K73 mutation not only disrupted this suppression but also fully inactivated the catalytic function. This contrasts with the partial effects of R35A/R38A/T915A, underscoring the unique role of K73 in both activity and regulation.

We next examined the physiological relevance of this observation by performing phage plaque assays with gradient-diluted T3 bacteriophage. Compared to OBCK (plaques visible after a 7-fold dilution), cells expressing OBC showed reduced plaque formation (plaques visible after a 6-fold dilution) (Fig. S5B). Furthermore, both OBCE and OBC exhibited the same plaque formation pattern, indicating that the wild-type OBC has partial phage resistance, a function that is abolished by the K73 mutation, while E255 has no significant effect on function. These in vivo experimental results are consistent with our biochemical data and collectively suggest that K73 is essential for both ATPase activity and antiviral function.

2.4. Potential drug targets and active sites of the protein

We performed an interaction interface analysis of the OBCK protein complex and compared it with the OBC protein complex. The binding sites changed from interactions between T442 and E31, E408 and E42, R35 and D416, and R38 and S36 to interactions between N5 and N21, F40 and V44, and T20 and N21. The non-covalent binding interactions at the interface of the OBCK protein complex were reduced (Figs. 5(a)–5(d)). Additionally, we found that the OBC protein complex interaction interface contains the E255 residue, which forms a hydrogen bond with H22, whereas the OBCK complex interaction interface does not contain the K73 residue. Changes in these sites may affect the function of the BREX defense system. In the Gabija system, which contains an ATPase domain, the top rim–ATPase dimer interface forms salt bridges, while there are fewer non-covalent interactions at the ATPase dimer interface. When key sites such as R296, K304, K360, K364, and K439 in the top rim–ATPase dimer interface, as well as R139 and K281 in the ATPase dimer interface, are mutated, the non-covalent interactions between sites such as R139 and K281 are disrupted, leading to the loss of the Gabija system protein's aggregation ability. This prevents tetramer formation, causing the Gabija system to lose its defense function.[19] Similarly, as a protein defense system containing an ATPase domain, mutation of key sites R38 and R35 in the OBC protein complex disrupts non-covalent interactions between them, thereby impairing the aggregation ability of the BREX defense system protein. We compared the OBC protein complex with the OBCE protein complex, as shown in Fig. 5(e). The comparison of the OBC protein complex with the OBCK protein complex is shown in Fig. 5(f). The difference between OBC and OBCK is attributable to the K73 mutation, which results in a root mean square deviation (RMSD) of 8.779 (Fig. S6A), indicating a substantial conformational change. In addition, the K73 mutation indirectly disrupts ATPase activity of the BREX defense system, thereby significantly impairing the normal function of its associated protein components (Fig. 5(f)).

Subsequently, we analyzed the electrostatic potential, structural flexibility, and thermodynamic stability of the wild-type OBC protein and its mutants. Electrostatic surface mapping revealed that the K73 mutation markedly reduced the positive potential within the ATP-binding pocket, likely weakening ATP binding, consistent with enzymatic assays showing a near-complete loss of ATPase activity (Fig. S7). Comparative flexibility analysis indicated that OBCK displayed reduced RMSF values in the Walker A motif (residues 67–74), suggesting increased rigidity that may hinder conformational changes required for ATP binding, whereas OBCE showed no significant change (Fig. S8). Thermodynamic stability predictions further indicated that OBCK is destabilizing ($\Delta\Delta G$: $-0.732/-0.289$ kcal/mol by mCSM/DUET), in contrast to the stabilizing effect of OBCE (0.404/0.769 kcal/mol), consistent with the lower yield and solubility observed for OBCK

during expression (Table S1). Together, these results suggest that K73 impairs ATP binding and catalytic activity by altering the electrostatic environment, reducing flexibility in a key motif, and destabilizing protein structure.

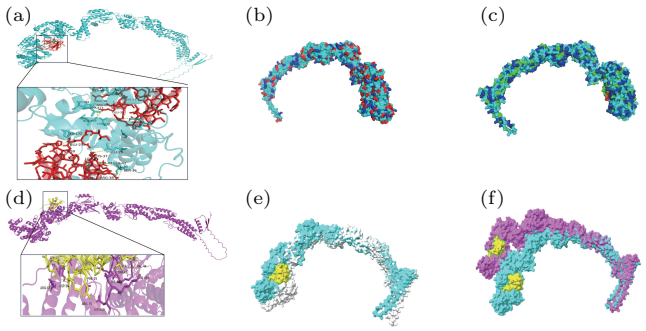


Fig. 5. Discovery of potential drug targets. (a) Interaction interface analysis of the OBC protein complex. (b) Distribution of hydrophobic and hydrophilic residues in the OBC protein complex. (c) Distribution of polar and nonpolar residues in the OBC protein complex. (d) Structural prediction and interaction interface analysis of the OBCK protein complex. (e) Structural comparison between the OBCE protein complex and the OBC protein complex; blue represents OBC, and white represents OBCE. (f) Structural comparison between the OBCK protein complex and the OBC protein complex; blue represents OBC, and purple represents OBCK.

To further explore the impact of the K73 mutation, we performed ATP molecular docking for OBC and OBCK using SwissDock (based on the EADock DSS engine) and visualized the results in PyMOL (Fig. S6B). The comparison revealed significant changes in the ATP-binding site after the K73 mutation. We then calculated the binding free energy (ΔG) between the protein and ATP using the PRODIGY-LIGAND tool, which employs a statistical potential model to assess the contribution of various atomic contacts at the protein-ligand interface. The calculation showed that the binding free energy for the OBCK mutant was -6.21 kcal/mol, slightly higher than the -6.54 kcal/mol for wild-type OBC, indicating that the mutation weakens ATP binding (Table S2). This change is primarily due to a significant reduction in polar interactions (e.g., CO and CN contacts), despite an increased number of nonpolar interactions (e.g., CC), which had a relatively limited effect on binding strength. These results collectively demonstrate the significant impact of the K73 mutation on conformational changes and the loss of ATPase activity.

In the OBC protein complex, R38 and R35 are analogous to the key aggregation sites in the Gabija defense system. These two sites may also be key aggregation sites in the BREX defense system, providing potential new targets for drug discovery against drug-resistant bacteria. Further analysis revealed that N398, R35, R38, and E31 are also B-cell epitopes, with R35 being particularly noteworthy. It is similar to the key sites R463 and E479 in the CRISPR-Cas3 defense system containing an ATPase domain, where these sites can activate CRISPR-Cas-related enzymes, affecting ATP depletion and inducing an immune response. Similarly, the R38 and E31 target sites in the BREX defense system also exhibit good immunogenicity, effectively activating the host's immune system and inducing specific immune responses, making them potential targets for subsequent vaccine development.

3. Methods

3.1. Plasmid synthesis

The BrxC protein sequence (NCBI: WP_000152592.1) and OrbA protein sequence (NCBI: WP_016033006.1) associated with the *Vibrio cholerae* strain were retrieved from the NCBI database. The pETDuet-N-6His-BrxC plasmid and pET28a-6His-OrbA plasmid were synthesized by Genescript Company in Nanjing.

3.2. Construction of the protein mutants

Based on the original plasmid, the BrxC-K73A and BrxC-E255A mutants were constructed. The primer sequences used in this study are shown in Table 1, and were synthesized by Beijing Ruixingko Biotechnology Co., Ltd.

Table 1. Primers used in this study.

Name	Sequence
BrxC-K73A-F	CgcaAGCTACTTCGGTAAAATGCTGGGTTACA
BrxC-K73A-R	TACCGAAGTAGCTtgcGCCGCTGCCGTAGAAGCC
BrxC-E255A-F	TTTTGATgcaGCGAGCGAGGCTATCTCCCAGA
BrxC-E255A-R	CGCTCGCtgcATCAAAAACAAAAACCAGAGTTTCC

3.3. PCR amplification of target genes

The synthesized BrxC plasmid served as a template for amplifying the target sequences of BrxC-K73A and BrxC-E255A. The PCR amplification system consisted of 0.5 μL template, 1.5 μL upstream primer (10 μM), 1.5 μL downstream primer (10 μM), 15 μL of 1 × PCR Mix, and ddH_2O up to 30 μL. The amplification reaction conditions were: predenaturation at 98 °C for 30 s, denaturation at 98 °C for 10 s, annealing at 60 °C for 30 s, and extension at 72 °C for 1 min, with 30 cycles, followed by a final extension at 72 °C for 1 min. The PCR products were digested with the restriction enzyme DpnI for 1 h. After digestion, the products were transformed into *E. coli* 3016, spread on LB solid medium containing ampicillin (Amp+), and incubated at 37 °C for 12 h. Colonies on the plate were picked for colony PCR identification. Colonies with successful PCR amplification were cultured in liquid LB medium and sent to Beijing Reborchen Biotechnology Co., Ltd. for sequencing.

3.4. Expression of target proteins

Referring to Gao *et al.*[31] with appropriate modifications, the synthetic pET28a-6His-OrbA plasmid and pETDuet-6His-BrxC plasmid or BrxC-K73A/E255A mutants were co-transformed into *E. coli* 3016 competent cells. Monoclonal

colonies were picked into 10 mL LB liquid medium with dual ampicillin resistance and cultured at 37 °C, 210 r/min for 8 h until the optical density at 600 nm (OD_{600}) reached 0.6. Then, 10 mL of the culture was expanded 1:100 into 1 L LB liquid medium and cultured at 37 °C, 200 r/min for 4–6 h until $OD_{600} = 0.6$. Subsequently, 0.25 mM IPTG was added to induce protein expression, and the culture was incubated at 18 °C for 12–16 h. The induced culture was centrifuged, and the resulting bacterial pellet was collected and stored at −40 °C for future use.

3.5. Purification of target proteins

As previously described,[32] the frozen bacterial pellet was thawed and resuspended in lysis buffer, then disrupted by sonication. The lysate was centrifuged at 20000 r/min and 4 °C for 1 h, and the supernatant was collected. A 20 μL aliquot of supernatant was mixed with 20 μL precipitation solution and 5 μL SDS loading buffer for protein denaturation, serving as the SDS-PAGE loading sample. The remaining supernatant was applied to a nickel-column for affinity purification. Non-specifically bound proteins were removed by washing with 20 column volumes of lysis buffer, and the target proteins were eluted with elution buffer. The eluted protein was further purified via ion-exchange chromatography on a heparin column (AKTA explorer). Fractions containing the target protein were collected and concentrated using 10 kD ultrafiltration tubes to obtain high-purity protein. Aliquots of 20 μL were reserved for subsequent SDS-PAGE analysis.

3.6. Modeling and optimization of tertiary structure

Based on the amino-acid sequence of the BrxC + OrbA complex, the tertiary structure of the BrxC + OrbA protein complex was predicted using AlphaFold 3 software. The model with the highest residue confidence score, as measured by the predicted local distance difference test (pLDDT), was selected for further evaluation. Structural quality was assessed using PROCHECK,[33] Verify3D,[34] and ERRAT.[35]

3.7. ATPase activity assay

ATPase activity was determined using the ATPase activity assay kit from Nanjing Jianchen Bioengineering Company. The OBC, OBCK, and OBCE complexes were first disrupted by ultrasonication, followed by centrifugation to collect the supernatant. The supernatant was then used for the activity measurement. The assay kit catalyzed the hydrolysis of ATP to ADP and inorganic phosphate (Pi). Subsequently, a phosphate indicator was added, which produced a blue color upon reaction with Pi. The absorbance was measured at 660 nm using a microplate reader, and the ATPase activity was calculated based on a standard curve derived from known phosphate concentrations.

4. Conclusion

Bacteria, archaea, and other microorganisms have evolved a range of complex and sophisticated immune defense systems to resist attacks from foreign invaders. The BREX system is one such defense mechanism. There are six types of the BREX defense system, with type I being the most common. In this study, we analyzed a large number of sites in the OBC protein complex predicted by the artificial intelligence model. We found that this complex does not have a transmembrane structure, nor does it contain a signal peptide. It is neither a membrane protein nor a secretory protein. The secondary structure of the protein complex is predominantly composed of α-helix (Hh), which accounts for approximately 66.06%, indicating the stability of the protein structure. Post-translational modifications (PTMs) are essential for cells to perform their normal physiological functions.[22–25] Phosphorylation is one of the most common modifications, and it can increase protein solubility, conformation, interactions, signaling, and degradation. These factors are crucial for the parasitism and reproduction of pathogens.[26,27] This study analyzed 135 phosphorylation sites in the OBC protein complex, and we found that these phosphorylation sites could inhibit the activity of the BREX defense system by interfering with post-translational modifications of the OBC protein complex. Using the artificial intelligence model, we simulated the OBC protein complex and analyzed its interaction interfaces. The resulting analysis, including the Ramachandran plot and Verify3D scores, revealed a high-quality model of the OBC protein complex (Fig. 2). Based on the artificial intelligence methods, this study provides a foundation for identifying the key functional sites of the OBC protein complex.

Using molecular cloning techniques, we constructed and expressed the OBC protein complex, and further purified it through ion exchange chromatography to obtain the target protein band, indicating the successful construction of the OBC protein complex for the first time (Fig. 3). Sequence analysis and extensive alignment revealed that lysine K73 may be a key site. Based on this, we constructed and expressed the OBCK and OBCE protein complexes. The ATPase activities of the OBC, OBCK, and OBCE protein complexes were measured, and we found that the ATPase activity of the OBCK protein complex was completely lost. This loss was solely due to the mutation of K73, which disrupted the ATPase domain, confirming our previous hypothesis and indicating that K73 is a crucial site in the OBC protein complex (Fig. 4(b)). K73 is the key ATP-binding site in the Walker A motif presented by artificial intelligence methods. The mutation of K73 demonstrates that the OBCK protein complex loses its ability to bind ATP during the enzyme reaction kinetics process, while the alteration of protein interaction sites in the OBCE protein complex

does not impair its ATPase activity. This suggests that the ATPase plays a role in the ATP binding process within the OBC protein complex. It is challenging to impact the defense system by slightly altering enzyme activity. For instance, Sun *et al.*[28] found that six mutations were required to completely abolish the NAD^+ enzyme activity of the SPARTA defense system, Gao *et al.*[29] found that three mutations were needed for the short prokaryotic Ago protein to completely lose its defense system activity, and Wang *et al.*[30] reported that at least two mutations were required to cause the loss of enzyme activity in the STAND defense system. In contrast, in this study, after extensive site analysis using artificial intelligence methods, a single mutation was enough to cause the complete loss of protease activity in the defense system, successfully achieving the efficient construction of the protein complex.

Using artificial intelligence methods for simulation, we discovered that the interaction between the two proteins in the OBC protein complex forms non-covalent bonds, which enhance the stability of the OBC protein complex (Fig. 5(a)). At the interaction interface of the OBC protein complex, we identified the E255 target site, where E255 interacts with H22 to form hydrogen bonds, stabilizing the complex. Interestingly, K73 residues were not found at the interaction interface of the OBC protein complex. E255 is located at the OBC interaction interface, and when the E255 residue is mutated, it may lead to significant changes in the interaction interface. However, in reality, the change in the interface is minimal. In contrast, K73 is not located at the OBC protein interaction interface. Although mutating K73 might not seem to affect the interface initially, it actually causes substantial changes, which indirectly supports the effectiveness and accuracy of using AI methods (ONS) to target K73.

Next, we analyzed the OBC, OBCK, and OBCE protein complexes using AI technology. We found that the interaction interface of the OBCK protein complex changed, with a reduction in non-covalent bonds, but new non-covalent bonds were also formed (Fig. 5(d)). BrxC, a homolog of BrxX, interacts with four residues of Ocr, forming hydrogen bonds with T82, N823, T768, and G721, and stabilizes Ocr through a salt bridge between R1145 and E67. Ocr induces the protein to adopt a closed conformation, which flips the MTase domain and anchors adenine, while the Y511A mutation disrupts the protein's specific and non-specific DNA binding.[15] As a homolog of BrxX, BrxC may similarly affect the internal protein interactions in the OBC protein complex, leading to the loss of BREX system defense functions. In the ATPase domain protein Gabija, a mutation at K78, which significantly reduces DNA binding and eliminates DNA cleavage, leads to the loss of the Gabija defense system's function.[19] Given that the OBC protein complex also contains an ATPase domain, it is possible that similar amino acid mutations could result in the loss of BREX system activity. By comparing the models of the two systems, we observed that the negative control OBCE protein complex is largely similar to the OBC protein complex (Fig. 5(e)). In contrast, the OBCK protein complex differs significantly from the OBC protein complex (Fig. 5(f)). This indicates that mutation of key sites can completely abolish ATPase activity and lead to significant changes in the overall structure of both complexes, resulting in alterations at critical target sites like R38 and R35. These findings suggest potential drug targets for further investigation.

Based on the key sites identified in the OBC protein complex related to the defense system, we carried out research on newly constructed proteins. We successfully constructed, expressed, and purified the OBC protein complex for the first time, and also purified the newly constructed OBCK and OBCE protein complexes. Further verification revealed that the ATPase activity of the OBCK protein complex was completely lost, which had a strong destructive impact on the BREX defense system. This single construction led to the complete disruption of the BREX defense system, proving that our method is not only feasible but also highly efficient. Additionally, using this efficient construction approach, we have preliminarily identified several potential drug targets and important antigenic epitopes, providing a theoretical foundation for the development of drugs against drug-resistant bacteria.

Acknowledgments

We thank X. Gao and S. Cui for providing technical support, and we thank the staff from the Core Facility of the National Institute of Pathogen Biology, Chinese Academy of Medical Sciences, and the Beijing National Laboratory for Condensed Matter Physics and the Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences. Special thanks to H. Zhu from IOP for valuable discussions and technical assistance. We are also grateful to H. Zhang, J. Han, and F. Liao from Tianjin Medical University for their assistance with this study.

This project was supported by the National Natural Science Foundation of China (Grant No. 82272308).

References

[1] Labrie S J, Samson J E and Moineau S 2010 *Nat. Rev. Microbiol.* **8** 317
[2] Deveau H, Garneau J E and Moineau S 2010 *Annu. Rev. Microbiol.* **64** 475
[3] Wang J Y, Pausch P and Doudna J A 2022 *Nat. Rev. Microbiol.* **20** 641
[4] Tock M R and Dryden D T F 2005 *Curr. Opin. Microbiol.* **8** 466
[5] Stokar-Avihail A, Fedorenko T, Hör J, Garb J, Leavitt A, Millman A, Shulman G, Wojtania N, Melamed S, Amitai G and Sorek R 2023 *Cell* **186** 1863
[6] Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, Afik S, Ofir G and Sorek R 2015 *EMBO J.* **34** 169
[7] Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J and Bernheim A 2022 *Nat. Commun.* **13** 2561

[8] Gordeeva J, Morozova N, Sierro N, Isaev A, Sinkunas T, Tsvetkova K, Matlashov M, Truncaite L, Morgan R D, Ivanov N V, Siksnys V, Zeng L and Severinov K 2019 *Nucleic Acids Res.* **47** 253

[9] Beck I N, Picton D M and Blower T R 2022 *Curr. Res. Struct. Biol.* **4** 211

[10] Shen B W, Doyle L A, Werther R, Westburg A A, Bies D P, Walter S I, Luyten Y A, Morgan R D, Stoddard B L and Kaiser B K 2023 *Nucleic Acids Res.* **51** 3513

[11] Zaworski J, Dagva O, Brandt J, Baum C, Ettwiller L, Fomenkov A and Raleigh E A 2022 *PLoS Genet.* **18** e1009943

[12] Isaev A, Drobiazko A, Sierro N, Gordeeva J, Yosef I, Qimron U, Ivanov N V and Severinov K 2020 *Nucleic Acids Res.* **48** 5397

[13] Andriianov A, Trigüis S, Drobiazko A, Sierro N, Ivanov N V, Selmer M, Severinov K and Isaev A 2023 *Cell Rep.* **42** 112972

[14] LeGault K N, Hays S G, Angermeyer A, McKitterick A C, Johura F T, Sultana M, Ahmed T, Alam M and Seed K D 2021 *Science* **373** eabg2166

[15] Li S, Xu T, Meng X, Yan Y, Zhou Y, Duan L, Tang Y, Zhu L and Sun L 2024 *Nucleic Acids Res.* **52** 8580

[16] Drobiazko A, Adams M C, Skutel M, Potekhina K, Kotovskaya O, Trofimova A, Matlashov M, Yatselenko D, Maxwell K L, Blower T R, Severinov K, Ghilarov D and Isaev A 2025 *Nat. Commun.* **16** 1825

[17] Studier F W and Movva N R 1976 *J. Virol.* **19** 136

[18] Oshiro R T, Dunham D T and Seed K D 2024 *J. Bacteriol.* **206** e0020624

[19] Li J, Cheng R, Wang Z, Yuan W, Xiao J, Zhao X, Du X, Xia S, Wang L, Zhu B and Wang L 2024 *Nature* **629** 467

[20] Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta R K and Kumar P 2021 *Mol. Divers.* **25** 1315

[21] White S R and Lauring B 2007 *Traffic* **8** 1657

[22] Li Y, Li Z, Yan P, Hua C, Kong J, Wu W, Cui Y, Duan Y, Li S, Li G, Ji S, Chen Y, Zhao Y, Yang P, Hu C, Lu M, Chen M and Xiao Y 2025 *Science* **387** eadr0393

[23] Kelly M, Dietz C, Kasson S, Zhang Y, Holtzman M J and Kim I K 2024 *Biochem. Biophys. Res. Commun.* **720** 150101

[24] Neagu A N, Josan C L, Jayaweera T M, Morrissiey H, Johnson K R and Darie C C 2024 *Molecules* **29** 4156

[25] Ryšlavá H, Doubnerová V, Kavan D and Vaněk O 2013 *J. Proteomics* **92** 80

[26] Liu J, Xu X, Zhong H, Yu M, Abuduaini N, Zhang S, Yang X and Feng B 2024 *Biomedicines* **12** 1446

[27] Cilek N, Ugurel E, Goksel E and Yalcin O 2024 *J. Cell. Physiol.* **239** e30958

[28] Sun D, Zhu K, Wang L, Mu Z, Wu K, Hua L, Qin B, Gao X, Wang Y and Cui S 2024 *Cell Rep.* **43** 114391

[29] Gao X, Shang K, Zhu K, Wang L, Mu Z, Fu X, Yu X, Qin B, Zhu H, Ding W and Cui S 2024 *Nature* **625** 822

[30] Wang Y, Tian Y, Yang X, Yu F and Zheng J 2025 *Nat. Commun.* **16** 2408

[31] Gao X, Zhu K, Qin B, Olieric V, Wang M and Cui S 2021 *Nat. Commun.* **12** 2843

[32] Gao X, Mu Z, Qin B, Sun Y and Cui S 2017 *Front. Cell. Infect. Microbiol.* **7** 411

[33] Laskowski R A, MacArthur M W, Moss D S and Thornton J M 1993 *J. Appl. Cryst.* **26** 283

[34] Lüthy R, Bowie J U and Eisenberg D 1992 *Nature* **356** 83

[35] Colovos C and Yeates T O 1993 *Protein Sci.* **2** 1511