Beyond the Active Site: The addition of a remote loop reveals a new complex biological function for chitinase enzymes.

Dan Kozome[1], Adnan Sljoka[2,3], Paola Laurino[1]*

[1] Protein Engineering and Evolution Unit, Okinawa Institute of Science and Technology Graduate University (OIST), Okinawa 904-0495, Japan.

[2] Center for Advanced Intelligence Project, RIKEN, Tokyo 103-0027, Japan

[3] Department of Chemistry, York University, Toronto, ON M3J 1P3, Canada

Corresponding author: Paola Laurino

Email: paola.laurino@oist.jp

Abstract

Loops are small secondary structural elements that play a crucial role in the emergence of new enzyme functions. However, our understanding of loop functions is mainly limited to the catalytic loops. To understand the function of remote loops in enzymes, we studied Glycoside hydrolase family 19 (GH19) chitinase - an essential enzyme family for pathogen degradation in plants. By revealing the evolutionary history and loops appearance of GH19 chitinase, we discovered that one loop which is remote from the catalytic site, is necessary to acquire the new antifungal activity. We demonstrated that this remote loop directly accesses the fungal cell wall, and surprisingly, it needs to adopt a defined structure supported by long-range intramolecular interactions to perform its function. Our findings prove that Nature applies this new strategy at the molecular level to achieve a complex biological function while maintaining the original activity one in the catalytic pocket, suggesting an alternative way to design new enzyme function.

Introduction

Protein evolution is an essential process that drives diversity in nature by enabling the emergence of new functions[1]. Understanding how protein evolves is fundamental, not only for unraveling the natural diversification of proteins but also for designing new protein functions *in vitro*. The structure of more than 30,000 proteins reveals that nature uses a limited number of basic structures (scaffolds) to achieve an enormous variety of functions[2]. Indeed, many reports demonstrate that this diversity is often acquired by diversifying the flexible catalytic loop structures exposed on the protein's surface, while preserving the robust scaffold and the active site residues common among the enzyme family as core structures[3–5]. For instance, altering catalytic specificity by grafting loop[6] or shaping the catalytic activity *via* mutations in the catalytic loop[7]. These strategies are a major source of structural and functional variation within protein superfamilies[8]. Thus, mutations in the catalytic loop regions have garnered significant attention because they can directly impact function. Indeed, because of their functional roles like ligand binding[9], promoting function via conformational changes[10], and functional switching[11], the propensity of loop regions to acquire mutations is evolutionarily advantageous. However, while previous studies have extensively explored the catalytic loops[6,71213] , the role of functional remote loop regions remains unexplored.

To understand how nature acquires new functions *via* remote loops, we investigated the evolution of Glycoside Hydrolase family 19 chitinases (GH19 chitinase; EC 3.2.1.14). GH19 chitinase hydrolyzes the glycosidic bonds of chitin, namely beta-1, 4-linked *N*-acetyl-D-glucosamine. Chitin is the main component of exoskeletons and cell walls of various organisms, including arthropods and fungi. Despite lacking an endogenous substrate for plant chitinases, many plants synthesize various chitinases. One of the physiological roles of chitinase is to

defend plants against pathogenic fungi by degrading chitin, a major component of the cell wall of many fungi[14]. Interestingly, even in GH19 chitinases, some chitinases do not exhibit antifungal activity. Taira *et al.* reported a difference in the antifungal activity of two chitinases from plants[15], - in this work, we will call them loopless and loopful GH19 chitinases. Comparing the sequences and the structures of both GH19, the catalytic residues and the core structure that consists of the catalytic cleft are conserved. However, loopful GH19, a chitinase from rye seed, contains six loop regions, whereas loopless GH19, a chitinase from moss, lacks five of those loop regions seen in loopful GH19 (Fig. 1A and B). Interestingly, loopful GH19 exhibits antifungal activity against the fungi *Trichoderma sp*., while loopless GH19 does not[15]. However, it is unclear how chitinase could achieve a gain in antifungal activity without any apparent effect on catalytic activity.

Herein we investigated the evolution of GH19 chitinase to understand how functional innovation occurred. We inferred five ancestral proteins with different loop combinations in remote loop regions. By characterizing these proteins variants, we discovered that loop II is critical for acquiring antifungal activity. Using structural analysis, molecular dynamics simulations, and computational studies, we showed that this remote loop allows new function through fine-tuned intramolecular interactions. Imaging studies via fluorescence microscope experiments unraveled that the acquisition of remote loop II access the substrate in a new specific cellular location, the fungal cell wall. This is the first report that highlights how Nature acquired remote loops in a protein structure to access a new and complex biological function.

Results

Phylogenetic analysis reveals the emergence of antifungal activity in GH19 chitinase. To identify the remote loop regions acquisition that played a role in the emergence of antifungal activity and when such an acquisition occurred in GH19 chitinase, we performed phylogenetic

analysis and ancestral sequence reconstruction. After extensive collection of all available sequences belonging to GH19 chitinase from plants and the removal of redundancy, we inferred the maximum-likelihood phylogenetic tree of 179 GH19 chitinase sequences. We used ancestral protein reconstruction to infer the most likely amino acid sequence for each ancestral node in the phylogeny. To avoid the bias that phylogenetic classification is affected by the presence or absence of loop regions, we inferred two phylogenetic trees using a multiple sequence alignment (MSA) of the collected sequences, and a MSA modified by removing the six loop regions in question. After we confirmed that the tree topology did not change based on the presence/absence of loop regions in the MSA (Fig. S2 and S3), we used the tree based on the unmodified MSA and selected five ancestral nodes where InDels of the loop regions occurred, designated Anc1 to Anc5 based on the presence/absence of loop regions in the extant sequences (Fig. 1C). We experimentally characterized chitinase activities and antifungal activities of five reconstructed ancestral proteins and two extant GH19 sequences. All five ancestral proteins showed a similar level of chitinase activity to the two modern GH19 chitinases (Table 1), while Anc5 also exhibited antifungal activity. Anc1-4 did not exhibit antifungal activity (Fig. 1D), suggesting that antifungal activity seems to be acquired during the transition from Anc4 to Anc5.

Characterization of the functional intermediates between Anc4 and Anc5. To identify the key loop insertion for antifungal activity acquisition, we constructed six variants of two reconstructed ancestral proteins, Anc4 and Anc5 with different combinations of loop I and II (Anc4 + Loop I, Anc4 + Loop II, Anc4 + Loop I and II, Anc5ΔLoop I, Anc5ΔLoop II, and Anc5ΔLoop I and II). We experimentally characterized antifungal activities and hydrolytic activities of these variants. To compare the strength of their antifungal activities, we calculated IC_{50} by performing a hyphal re-extension inhibition assay. Addition/removal of loop I and/or II regions did not affect their hydrolytic activities (1.1 – 1.5-fold) except for in the case of loop

I addition to Anc4. Insertion of loop I to Anc4 resulted in a loss of catalytic activity of Anc4 (Fig. 2B). However, removing loop II from Anc5 decreased its antifungal activity by 12-fold while removing loop I did not influence its antifungal activity (roughly 1.02-fold increase, Fig. 2B). These results strongly suggest that loop II has a role in enhancing antifungal activity. However, it is worth noting that the addition of loop II to Anc4 did not improve its antifungal activity. Anc4 and 5 differ by 45 substitutions (Fig. 2A, C), suggesting that loop II enhances antifungal activity in combination with residues in the protein scaffold.

Structural analysis of Anc4 and Anc5 and molecular dynamics (MD) simulations revealed the importance of long-range interactions in the emergence of antifungal activity.

To get structural insights into the emergence of antifungal activity in GH19 chitinase, we first solved the X-ray crystal structure of Anc4 and Anc5 (Table S3). Overall, the backbone structures of Anc4 and Anc5 were nearly identical (RMSD of Cα = 0.478 Å) and the orientation of catalytic residues is similar, suggesting that loop I and II addition/removal did not cause major structural disruptions. This is supported by the fact that addition/removal of loop I and/or II regions did not affect their hydrolytic activities except for the loop I insertion into Anc4. The loss of the hydrolytic activity in Anc4 via loop I insertion might be due to the lack of disulfide bonding between loop I and the core scaffold. This is because the cysteine residue that forms the disulfide bonding with loop I is replaced by another residue (Val 81 in the MSA, Fig. S5). Although Anc4 and Anc5 showed no significant structural difference and their mutants Anc4 + Loop II and Anc5ΔLoop I had similar hydrolytic activity, their antifungal activities were significantly different (Anc5ΔLoop I showed 56-fold higher activity than Anc4 + Loop II). Therefore, we explored the contributions of protein dynamics to the acquisition of antifungal activity by performing molecular dynamics (MD) simulations. We performed four runs of 200 ns simulations of the model structure of Anc4 + Loop II and Anc5ΔLoop I built using Anc4

and Anc5 structures (PDB 8HNE and 8HNF). In Anc5ΔLoop I, the simulation showed reduced mobility of the loop II regions and a loop region between the ninth and tenth α-helices (positions 192 - 201) compared to Anc4 + Loop II (Fig. 3A, B). Some of the 45 substitutions support the stabilization of loop II in Anc5ΔLoop I. Structural comparison of Anc4 and Anc5 revealed six substitutions forming new interactions with loop II regions (Fig. 3C). Two substitutions, P12K and N13H, formed new hydrogen bonding with Asp residue in loop II. The His 13 residue is also involved in hydrophobic interaction with Trp residue in loop II (Fig. 3C). D197(217)R (residue numbers in Anc5 state are in parentheses) substitution formed new hydrogen bonds between the N terminus and the tenth α-helix (Fig. 3C). N193(213)G substitution reduced structural hindrance with the Tyr residue in loop II (Fig. 3C). Loop II insertion into Anc4 caused a change in conformational orientation of Ser 58, leading to a formation of a hydrogen bond between the oxygen atom of Ser 58 and the nitrogen atom of the Gly 60 as seen between Thr 65 and Gly 67 in Anc5. Y194(214)F substitution reduced structural hindrance with the oxygen atom of Thr 65 (Fig. 3C). These substitutions are important for stabilizing the loop II region to perform antifungal activity.

To assess whether residues in the scaffold have long-range effects on rigidity and conformational modifications of loop II, we utilized rigidity transmission allostery (RTA) algorithms[16]. RTA is a computational approach based on mathematical rigidity theory[17,18] and graph theory, which analyzes long-range communication and allosteric networks within protein structures[19–22]. RTA measures whether local mechanical perturbation of rigidity at one site propagates and modifies rigidity and conformational degrees of freedom at distant site(s) in protein structure. Starting with a structure, RTA first utilizes the method Floppy Inclusion and Rigid Substructure Topography (FIRST)[23] to generate a constraint network, where protein structure is modeled in terms of vertices (atoms) and edges (i.e., covalent bonds, electrostatic bonds, hydrogen bonds, and hydrophobic contacts). Every potential hydrogen bond is ranked

and assigned energy strength according to its donor-hydrogen-acceptor geometry. Upon rigidification of individual site(s) (i.e. residues), RTA then quantifies transmission (changes) of degrees of freedom and strength of communication across protein structure. RTA analysis on Anc5ΔLoop I showed that several residues in the scaffold are involved in long-range communication with loop II, revealing an allosteric network of residues that transmit communication with loop II (Fig. 3D). Interestingly, many of the substitution residues are part of this communication network, suggesting they impact the stability and conformational dynamics of loop II. To further probe this, we applied FIRST and decomposed the protein structures into rigid and flexible regions. FIRST rigid cluster decomposition on Anc4 + Loop II and Anc5ΔLoop I structures predicted that loop II is stabilized by intra-hydrogen bonding between the Pro 72 and the Asp 73 (residue number is based on the MSA in Fig. S5A), which is observed only in Anc5ΔLoop I structure (Fig. S5C). This results in a strong rigid cluster in Anc5ΔLoop I which persists over wide hydrogen bond energy strengths (Fig. S5D, E).

Loop II leads to gain of antifungal activity by promoting binding to the fungal cell wall. To understand the complex biological role of loop II, we performed fluorescence microscope experiments. These experiments aimed at verifying if the acquisition of loop II in GH19 chitinase plays a role in binding to the surface of the fungal cell wall. For this purpose, we prepared the catalytically inactive mutants of Anc4, Anc4 + Loop II, Anc5, Anc5ΔLoop I, Anc5ΔLoop II, and Anc5ΔLoop I and II by replacing catalytic glutamate residue with glutamine (The Glu 67 in the MSA, Fig. S2). This mutation makes proteins lose chitin degrading activity in the fungal cell wall. These inactive mutants were tagged with AlexaFluor488. By fluorescence microscope experiments, we observed that only Anc5 and Anc5ΔLoop I bound to the surface of fungal hyphae (Figs. 4 and S6). This result is consistent with the result that Anc5 and Anc5ΔLoop I showed 28-fold stronger antifungal activity than

Anc4 (Fig. 2b), suggesting that the acquisition of loop II is necessary to access the substrate in the cell wall of fungi to perform antifungal activity.

Discussion

Proteins use the limited numbers of folds and have evolved their structure and function using substitutions and insertions and deletions (InDels) of loops and other secondary structures[24,25]. However, most studies focused on substitutions and InDels are rarely considered. In these rare cases, when InDels are studied, they consist of one to three amino acid removal or addition[26,27] or graft of the catalytic loops[6,7,12]. Shining light on the roles of insertion or deletions of entire remote loop regions in protein evolution can provide further understanding of how we can engineer new protein functions. Our study reveals the molecular mechanism of an enzyme, through addition of a remote loop, can acquire a new distinct function distant from the catalytic site while maintaining its original activity.

To address this, we first identified the transition where the remote loop regions were inserted or deleted during evolution of the enzyme. This is the first study, to out best knowledge, that demonstrates that phylogenetic analysis and ancestral sequence reconstruction can explore remote loop acquisition in enzymes and thus identify potential hotspots for loop engineering. Most studies using ancestral sequence reconstruction have been performed with a fixed length of multiple sequence alignment due to the ambiguous evolutionary information of InDels. In our case, the function of GH19 chitinase varies from chitinase activity to antifungal activity, depending on the presence/absence of remote loop regions. Our phylogenetic analysis with two multiple sequence alignments (with/without the considered loop regions) showed that InDels of remote loop regions did not occur frequently and the tree topology did not change

significantly due to the InDels of remote loop regions (Fig. S3). Thus, we identified the ancestral nodes where the protein acquired/lost remote loops and inferred functional ancestral proteins with different loop combinations (InDels). As previous studies[27–29] demonstrated that InDels in loop regions are highly tolerated, our ancestral proteins showed robustness to add/remove loops. Thus, we demonstrate that ancestral sequence reconstruction approach in enzyme family with structural variations in loop regions is useful for designing enzyme with different loop combinations and exploring the potential remote site for loop engineering.

Surprisingly these ancestral proteins have the same sequence length and almost identical structure (RMSD of $C\alpha = 0.478$ Å), however, they showed significantly different antifungal activity (Anc4 + Loop II and Anc5 ΔLoop I), indicating that structural dynamics play an important role for the additional functions[30]. In the evolution of GH19 chitinase, stabilizing the key remote loop regions is important to perform new protein functions, and this stabilization requires additional substitutions. Even if the order in which InDels or substitutions occurred is not a trivial task, previous studies suggest that drastic mutational steps occur in combination with substitutions[6,7,31].

In the evolution of GH19, it is interesting to note that the enzyme became dual-functional with a new complex biological activity while maintaining its original activity even though a trade-off between the catalytic activities to original substrate and the promiscuous substrate is often observed in protein evolution[3233]. Trade-off can be associated with a change in the conformational dynamics of the catalytic pocket optimizing to the substrate[30], and bifunctionality can be achieved when this trade-off is weak enough to maintain original activity[34]. In the case of GH19, there is no significant change in the dynamics near the catalytic residues due to the mutations and original catalytic activities retained (Figs. 2B, 3A, and S5).

Thus, the enzyme seems to have developed a dual-function rather than expanding its promiscuous activity.

Strikingly, our findings show that the key loop insertion allowed the accessibility to the substrate in a different cell types and cellular location (Fig. 4). Since the target substrate for antifungal activity is located in the cell wall of fungi, not in the solution, the protein needs to access the substrate location to perform protein function. The fluorescence labeling experiments clearly show that chitinase access to the fungal cell wall to acquire antifungal activity (Figs. 2B, 4, and S6). Nevertheless, the acquisition of this new and complex function is possible only through the long-range interaction between the inserted secondary element and the substituted residues in the protein scaffold (Figs. 3 and S5). Herein we are the first to report how remote loop insertion plays an important role in accessing the substrate.

In addition, our findings are key to develop protein engineering methods aimed at accessing water-insoluble substrates such as cellulose, chitin, and plastic. To achieve efficient degradation of water-insoluble substrates, accessibility is equally important to catalytic efficiency[35]. On the other hand, here, we are the first to prove how functional peptide grafting is not only limited to catalytic loop, but it can be extended to remote loops. In conclusion, we show an alternative and effective way to redesign or expand enzyme function, opening new ways of thinking for enzyme designers.

Methods

Phylogenetic analysis and ancestral sequence reconstruction

2617 sequences of plant GH19 chitinase were collected from the UniProtKB database[36]. Redundant sequences with more than 50% sequence identity were filtered using the CD-HIT

program. The resulting 682 sequences were aligned using MAFFT ver. 7[37]. Only GH19 catalytic domain sequences were aligned, and additional domains were manually removed using an alignment visualizing software, Aliview[38]. Sequences of a loopless type GH19 chitinase from *Gemmabryum coronatum* (Uniprot: A9ZSX9; residues 25-228) and a loopful type GH19 chitinase from *Secale cereale* (Uniprot: Q9FRV0; residues 24-266) were used as references for the smallest and largest GH19 catalytic domain, respectively. 11 bacterial GH19 chitinase sequences were added to the dataset as an outgroup. The resulting 179 sequences were aligned with MAFFT and a maximum-likelihood phylogenetic tree was estimated using the model (WAG+F+I+G4) automatically determined in IQ-TREE[39]. Ancestral protein sequences were reconstructed using the empirical Bayesian method applied by IQ-TREE[35]. The ancestral sequences Anc1 to Anc5 were reconstructed using the WAG substitution matrix together with the maximum-likelihood phylogenetic tree.

Cloning and site-directed mutagenesis

Codon-optimized genes encoding the ancestral GH19 chitinase proteins and loopful type GH19 chitinase from *Secale cereale* (Uniprot: Q9FRV0; residues 24-266) were synthesized by TWIST Bioscience and cloned into the pET-22b (+) vector using the iVEC3[40]. PCR amplifications for synthetic genes and a linear-pET-22b (+) vector were performed with using PrimeSTAR® Max DNA polymerase (TaKaRa) and the designed primers (Supplementary Table 1) containing appropriate overlapping regions for iVEC3. The gene coding loopless type GH19 chitinase from *Gemmabryum coronatum* (Uniprot: A9ZSX9; residues 25-228) cloned into the pET-22b (+) vector was a gift from Toki Taira.

Site-directed mutagenesis was achieved by Inverse PCR using PrimeSTAR® Max DNA polymerase (TaKaRa) with the designed primers (Supplementary Table 1). Successful cloning and mutagenesis were confirmed by Sanger sequencing.

Protein expression and purification

E. coli SHuffle T7 (DE3) cells harboring the protein gene of interest were grown in LB at 37 °C to OD_{600} 0.6-0.8, induced with 0.1 mM β-d-1-isopropyl thiogalactopyranoside (IPTG) and incubated further for 24 hours at 18 °C. Cells were pelleted and stored at −80 °C before protein purification. The cells were disrupted by sonication in a 20 mM Tris-HCl buffer, pH 8.0. The sonicated extract was separated into soluble and insoluble fractions by centrifugation at 12,000×g for 15 minutes at 4 °C. The soluble fraction was dialyzed against 10 mM sodium acetate buffer, pH 5.0, and filtered before applying to a RESOURCE Q column (6 mL, Cytiva) or HiTrap SP HP column (5 mL, Cytiva) previously equilibrated with the same buffer. The elution was done with a linear gradient of NaCl from 0 to 0.3 M in the same buffer. The recombinant protein fractions were collected and dialyzed against a 5 mM Tris-HCl buffer containing 150 mM NaCl, pH 8.0. Purified recombinant proteins were concentrated using a Millipore centrifugal protein concentration device (10 kDa cutoff) and loaded onto a Superdex200 Hiload 16/600 column (Cytiva) equilibrated with 5 mM Tris-HCl buffer containing 150 mM NaCl, pH 8.0. Protein purity was confirmed by SDS–PAGE, and protein concentrations were measured spectrophotometrically using molar absorption coefficients calculated in ProtParam (http://expasy.org/tools/protparam.html).

Chitinase activity assay

Chitinase activity was measured colorimetrically with glycol chitin as a substrate. Ten μL of the sample solution was added to 150 μL of 0.2% (w/v) glycol chitin solution in 0.1 M sodium acetate buffer, pH 5.0. After incubation of the reaction mixture at 37°C for 15 minutes, the reducing power of the mixture was measured with ferric ferrocyanide reagent by Imoto &

Yagishita[41]. One unit of activity was defined as the enzymatic activity that produced 1 mmol of GlcNAc per minute at 37°C.

Antifungal activity assay

Qualitative assay

An antifungal assay was performed according to the method of Schlumbaum *et al.*[14] with modification. An agar disk (4 mm in diameter) containing the test fungus, *T. longibrachiatum*, prepared from the cultured fungus on potato dextrose broth with 1.5% (w/v) agar (PDA), was placed in the center of a Petri dish containing PDA. Wells were subsequently punched into the agar at a 15 mm distance from the center of the Petri dish. 500 pmol of each protein sample was placed into the wells. The plates were incubated for 24 hours at 25°C.

Quantitative assay

Hyphal re-extension inhibition assay was done by using *T. longibrachiatum*. Agar disks (4 mm and 1 mm in diameter and in-depth, respectively) containing the fungal hyphae, which were derived from the resting part of the fungus previously cultured on potato dextrose broth containing 1.5% (w/v) agar (PDA), were put on another PDA plate with the hyphae attached side down. Five μL of sterile water or sample solution were overlaid on the agar disks, and then the plate was incubated at 25 ℃ for 12 hours. After incubation, images of the plates were scanned using an image scanner. The areas of the re-extended hyphae were calculated as numbers of pixels by GNU Image Manipulation Program (GIMP, ver. 2.0). The protein concentration required for inhibiting the growth of the fungus by 50% was defined as IC_{50} and determined by constructing dose-response curves (percentage of growth inhibition versus protein concentration).

Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) experiments were performed using StepOnePlus Instrument, a real-time PCR equipment. Reaction mixtures contained 2.5 μM protein in 10 mM sodium acetate buffer, pH 5.0, 5× SYPRO orange dye in a total volume of 20 μL and dispensed into a 96-well PCR plate. Fluorescence intensities were monitored continuously as the sample mixtures were heated from 20 ℃ to 99 ℃ at a rate of 1% (approximatelu 1.33 ℃/min), using the ROX channel. Melting temperatures (T_m) were determined by fitting the fluorescence derivative data to a quadratic equation in the vicinity of the Tm in R software.

Crystallization.

After purification, Anc4 and Anc5 were concentrated at 10 mg/ml and 5.67 mg/ml in the 5 mM Tris-HCl buffer, pH 8.0, respectively. Initial crystallization screens were performed using various crystallization screening kits commercially available. The protein solution drop (0.4 mL) was mixed with 0.4 mL of a reservoir solution and then equilibrated with 50 mL of the reservoir solution using a crystallization robot, Mosquito® Xtal3. The crystallization conditions were screened using the sparse-matrix sampling method, according to the sitting-drop vapor diffusion method at 20 ℃ in a 96-well plate (violamo). After a week, well-formed crystals of Anc4 and Anc5 were obtained using PEGRx 1 (Hampton Research) and further optimized to the condition (15% and 27% (w/v) polyethylene glycol monomethyl ether 2,000, 0.1 M MES monohydrate pH 6.0 for Anc4 and Anc5, respectively).

X-ray data collection

For data collection, the crystals of Anc4 and Anc5 were soaked in cryoprotectant buffer (40% (w/v) polyethylene glycol monomethyl ether 2,000, 0.1 M MES monohydrate pH 6.0) for 1 min before flash-cooling to 100K in liquid nitrogen. Diffraction datasets were collected at 100 K on BL32XU or BL41XU beamline of the SPring-8 (Harima, Japan), employing the

automated data collection ZOO[42]. The collected datasets were processed automatically with KAMO[43]. Each dataset was indexed and integrated using XDS[44], followed by a hierarchical clustering analysis using the correlation coefficients of the normalized structure amplitudes between datasets. Finally, a group of outlier-rejected datasets was scaled and merged using XSCALE[45] or Aimless[46].

Structure solution and refinement

General data handling was performed with the Phenix package[47]. The initial model was solved by molecular replacement using Phaser[48] with the model structure of Anc4 and Anc5 predicted by AlphaFold2[49] . The model building was performed with *Coot*[50]. Structures were refined by *Coot*[50] and Phaser[44]. The details of crystallization, structure determination, the data collection, data processing, and refinement statistics are given in Supplementary Tables 2 and 3. Structural figures are described and rendered by the PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

Molecular Dynamics Simulations

The MD simulations were performed using GROMACS version 2020.1[51] and the charmm36-mar2019 force field[52]. The model structure of Anc4+LII and Anc5ΔLI were built using Anc4 and Anc5 structures obtained from this work and used as a starting point for MD simulations. For each model, the system was solvated and neutralized with Na^+ and Cl^- ions in a dodecahedral box. Temperature was maintained at 310 K by using a modified Berendsen thermostat[53] and pressure was maintained at 1 bar by using a Parrinello-Rahman barostat[54]. The system was then equilibrated on energy, temperature, and pressure before performing four individual repeats of 200 ns of the simulations. To analyze the fluctuation, Gromacs standard analysis package was used.

Rigidity-based allosteric communication and FIRST

Allostery analysis was carried out by applying rigidity-transmission allostery (RTA) analysis[16] . The RTA method utilizes graph and rigidity theory[17,18] techniques to identify allosteric networks within structures of proteins and protein complexes[19–22]. Starting with protein structures of Anc4 + Loop II and Anc5ΔLoopI, we applied the RTA algorithm by sequentially perturbing rigidity of individual residues and monitoring changes in conformational degrees of freedom in loop II. Rigid cluster decomposition and dilution plots on Anc4 + Loop II and Anc5ΔLoopI were performed with software FIRST[23]. FIRST creates a geometric molecular framework, whose underlying network (graph) contains atoms (vertices) and edges (*i.e.* constraints representing covalent bonds, hydrogen bonds, electrostatic interactions, and hydrophobic contacts). Every potential hydrogen bond is assigned an energy strength in kcal/mol, and a hydrogen bond cutoff energy value was selected so that all bonds weaker than this cutoff are removed from the network. FIRST then applies the pebble game algorithm[18,55] which rapidly decomposes a protein structure into flexible and rigid regions while incrementally removing weak hydrogen bonds.

Microscopic observations

10 mL of sterilized PDB (Potato dextrose broth) medium containing 10^6 spores/mL of T.*longibrachiatum* was incubated at 25℃ for 24 h with shaking at 300 rpm to obtain the mycelia of the fungi. The mycelia were collected by centrifugation (3000*g*, 25 °C, 20 min) and was resuspended by PDB medium. Then, 50 μL of the mycelial suspension was mixed with 50 μL of 2 μM each protein sample solution and incubated at 25 °C for 1 h. The incubated solution mixture was washed three times with 20 mM sodium phosphate buffer (pH 7.4). The samples

were observed with confocal laser scanning microscopy. Each protein sample was tagged with Alexa Fluor 488 Microscale Protein Labeling Kit, followed by the manufacturer's protocol.

Figures

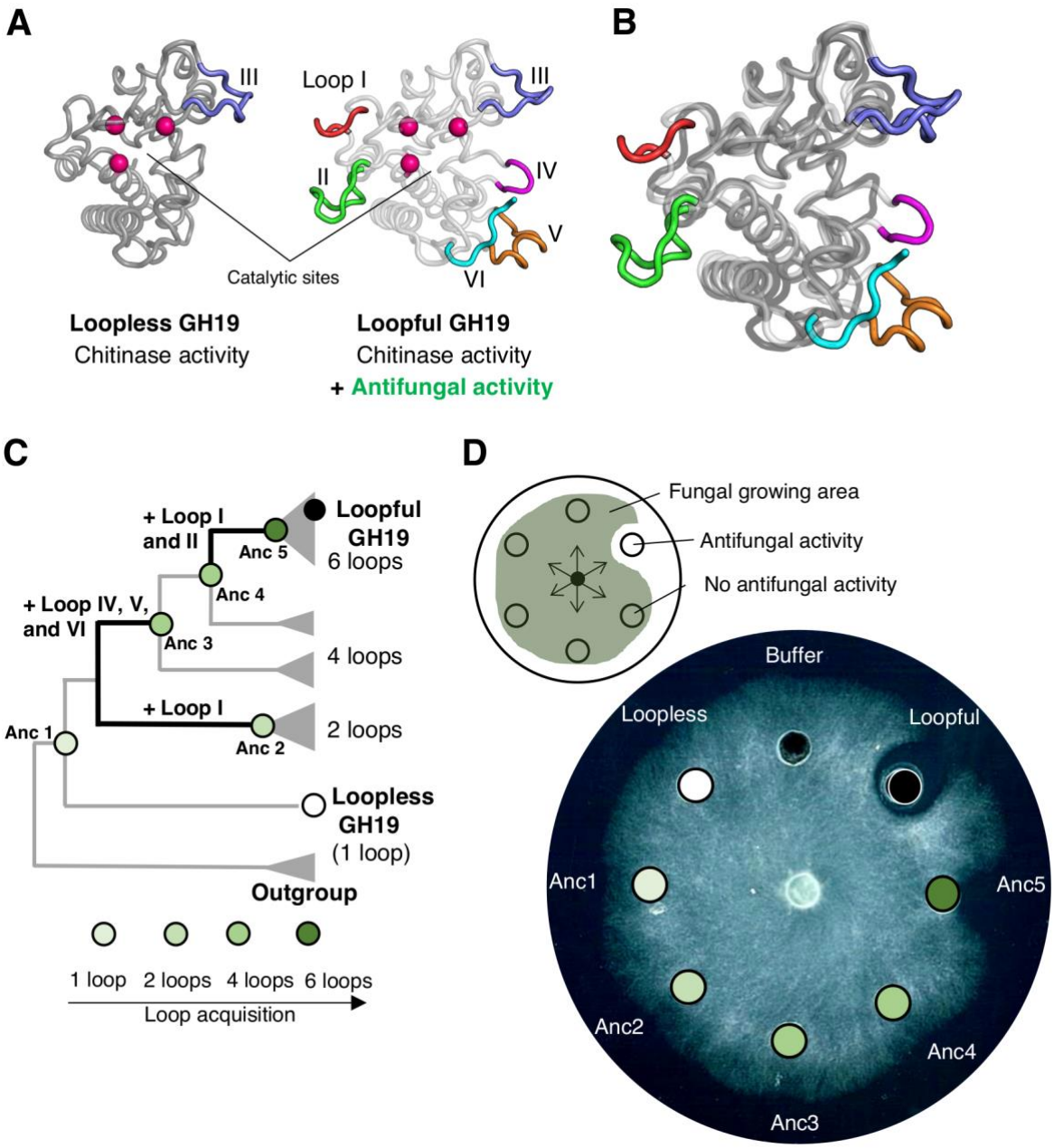
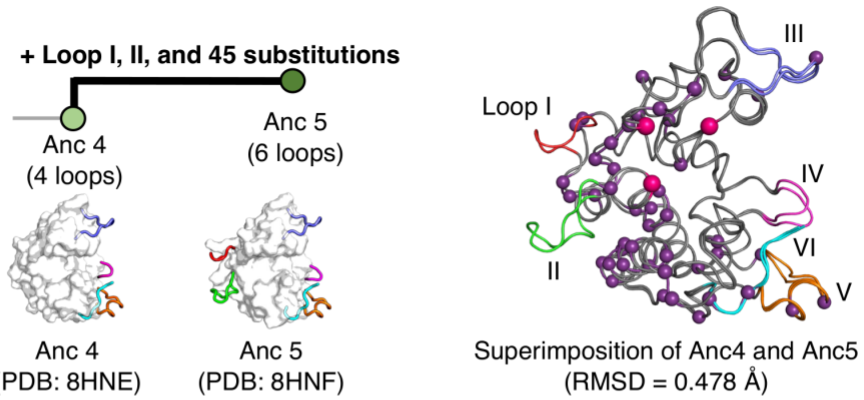


Fig. 1. Antifungal activity is acquired during evolution in GH19 chitinase. A) Structural comparison of two types of GH19 chitinase, loopless and loopful. They share the core region

(darker and lighter grey in loopless and loopful GH19, respectively), two catalytic glutamic acids, and one serine that holds catalytic water (magenta spheres) but differs in loop regions, named I to VI. (shown in red, green, light blue, purple, orange, and cyan, respectively). B) Structural superimposition of loopless and loopful. The detail catalytic mechanism that they conserve is shown in Fig. S1. **C**) Schematic representation of a phylogenetic tree of GH19 chitinase. The five ancestral nodes that were characterized are labeled and colored according to the number of loop regions (80% lighter green, one loop; 60% lighter green, two loops; 40% lighter green, four loops; 25% darker green, six loops). Evolutionary transitions containing loop acquisitions are highlighted with thick black lines. A multiple sequence alignment of two modern and five ancestor sequences and a full phylogenetic tree are shown in Figs. S2 and S3, respectively. **D)** Top, a schematic representation of fungal hyphae expansion inhibition assay against *Trichoderma longibrachiatum* as the test fungus. Bottom, the assay plate. Each well contains 10 μL of sterilized buffer (10 mM sodium acetate buffer, pH 5.0) or 500 pmol of protein samples dissolved in the same buffer.

A



B

Variants	Chitinase activity (U/mol) $\times 10^9$	Antifungal activity (IC_{50}, μM)	Thermostability (Tm, °C)
Anc 4	1.05	488 ± 3	72.5
Anc 4 + Loop I	N. D.	N. D.	64.0
Anc 4 + Loop II	1.15	737 ± 7	64.0
Anc 4 + Loop I and II	N. D.	N. D.	55.0
Anc 5	0.85	16.9 ± 1.4	66.0
Anc 5 Δ Loop I	1.28	17.2 ± 1.9	60.0
Anc 5 Δ Loop II	1.04	200 ± 20	62.0
Anc 5 Δ Loop I and II	1.28	575 ± 4	72.0

C

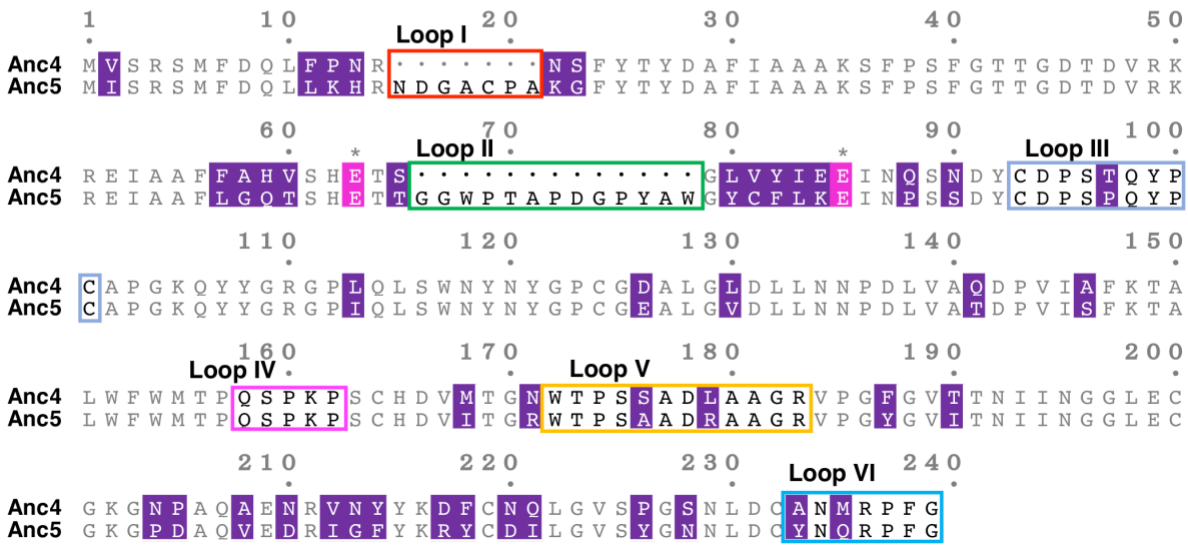


Fig. 2. Characterization of eight variants of two ancestral proteins reveal the key remote loop for the acquisition of antifungal activity. A) Left, a schematic representation of an evolutionary transition from Anc4 to Anc5. Crystal structures of Anc4 and Anc5 solved to 1.13

Å and 1.57 Å, respectively. Right, structural superimposition of Anc4 and Anc5. Two catalytic glutamic acids and one serine reside that holds catalytic water are indicated as magenta spheres. 45 substitution residues are indicated as purple spheres. **B**) Summary of the effects of loop presence/absence on the hydrolytic and antifungal activities. One unit of activity is defined as the enzyme activity that produced one μmol of GlcNAc per minute at 37°C. Error bars represent SD (n = 3). Melting temperature (T_m) was measured using differential scanning fluorimetry followed by the procedures as described in the Methods. **C**) An alignment of Anc4 + Loop II and Anc5ΔLoop I sequence. Loop regions II to VI are highlighted in green, light blue, purple, orange, and cyan squares, respectively. Two glutamate residues and one serine residue are shown in magenta. 45 substitutions residues are shown in purple.

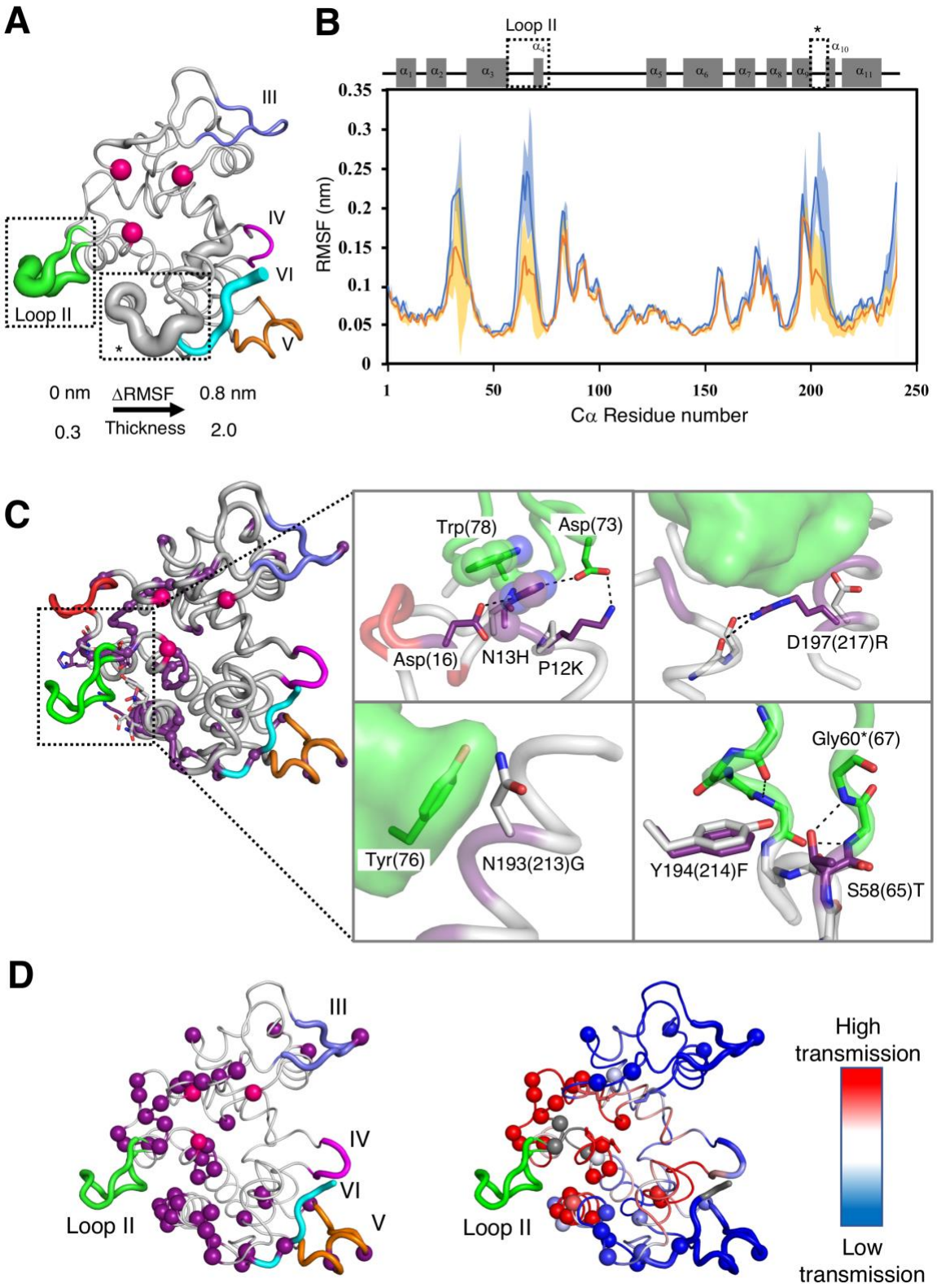


Fig. 3. **Comparison of Anc4 + Loop II and Anc5ΔLoop I exhibit the structural and dynamics contribution of Loop II and scaffold substitutions. A**) The difference in Δroot

mean square fluctuation (ΔRSMF = RMSF of Anc4 + Loop II – RMSF of Anc5ΔLoop I) is mapped on the structure of Anc5ΔLoop I as the thickness of the cartoon representation. Spheres indicates two catalytic glutamic acids and one serine that holds catalytic water. **B**) Plots of the RMSF of Cα of each residue in Anc4 + Loop II (blue) and Anc5ΔLoop I (orange). Error bars are indicated as shades. A schematic representation of secondary structures of GH19 chitinase is shown. Dashed squares indicate the region where ΔRSMF is more than 0.05 nm **C**) Cartoon_tube representation of Anc5. Sphere representation indicates 45 substitution residues (purple) and two catalytic glutamic acids and one serine that holds catalytic water (magenta). Resides around 4 Å from loop II are shown in sticks. Intra molecular interactions stabilizing loop II regions are shown in the enlargements. Resides in Anc4 and Anc5 state are shown in grey and purple sticks, respectively. Black dashed lines indicate hydrogen bonds. Residue numbers are based on Anc4. In the case that the number shifts due to loop insertions, residue numbers of Anc5 state are in parentheses. An asterisk indicates the number of Anc4 + Loop II. Loop regions I to VI are shown in red, green, light blue, purple, orange, and cyan, respectively. **D**) Left, computational analysis of long-range communication in Anc5ΔLoop I. Cartoon representation of Anc5 ΔLoop I with cartoon_tube representation of loop II to VI (shown in green, slate, purple, orange, and cyan, respectively). Sphere representation indicates 45 substitutions residues (purple) and two glutamic acids and one serine that holds catalytic water (magenta). Right, rigidity transmission allostery (RTA) communication analysis on Anc5ΔLoop I. Residues are colored based on the intensity (red being highest) of long-range rigidity transmission communication with loop II (green). Spheres indicate 45 substitutions.

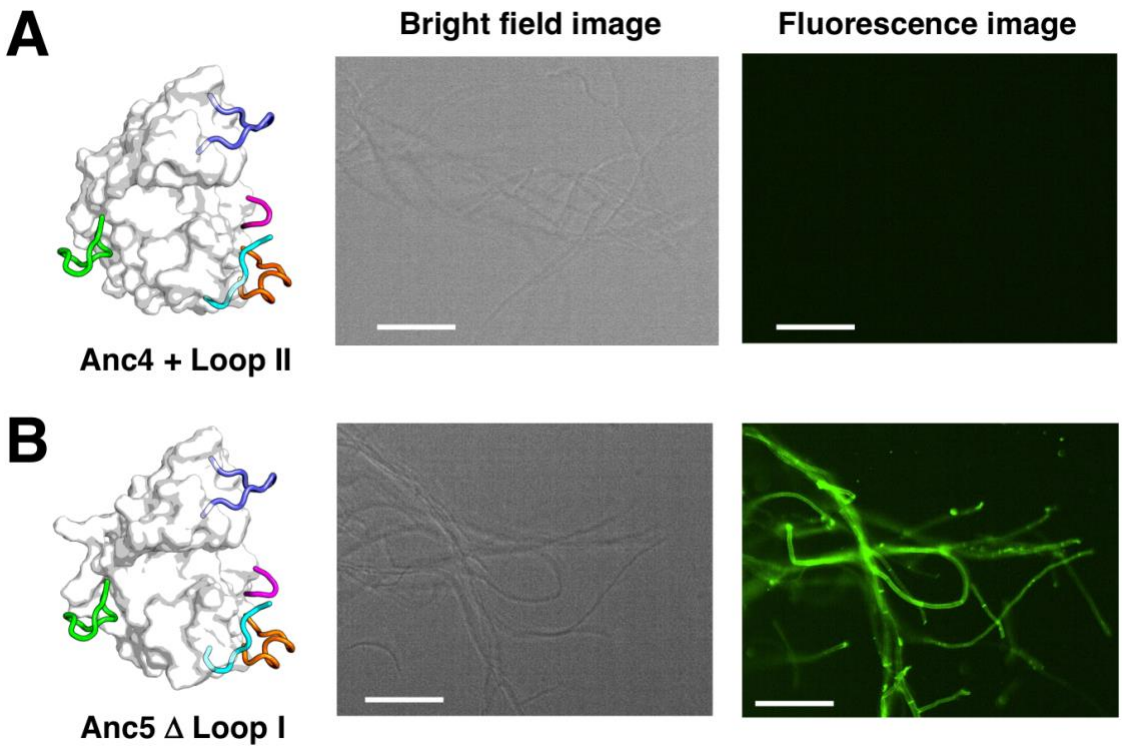


Fig. 4. Fluorescence microscopy images of fluorophore-labelled protein Anc4 + Loop II and Anc5ΔLoop I on the surface of fungal hyphae reveal the mechanism of acquisition of antifungal activity.

Each 50 μL of 2 μM Alexa Fluor 488-labelled proteins Anc4 + Loop II and Anc5ΔLoop I is mixed with *T. longibrachiatum* hyphae in 20 mM sodium phosphate buffer, pH 7.4 at 25 °C. Images were captured after washing excess of the labelled proteins. All scale bares are 100 μm. A) From left to right: the structure of Anc4 + Loop II; Bright field image of fungal hyphae; Fluorescence microscope image shows no binding to the surface of fungal hyphae. B) From left to right: the structure of Anc5 + Loop II; Bright field image of fungal hyphae; Fluorescence microscope image shows Alexa Fluor 488-labelled Anc5 ΔLoop I binding to the surface of fungal hyphae.

Table 1 Enzymatic activities and melting temperature of the two modern and five ancestral GH19 chitinases

Protein	Loop	Specific activity (U*/mol) × 10^9	DSF T_m (°C)
Anc1	III	1.44	58.5
Anc2	I and III	0.75	52.0
Anc3	III, IV, V, and VI	1.20	72.0
Anc4	III, IV, V, and VI	1.05	72.5
Anc5	I, II, III, IV, V, and VI	0.85	66.0
Loopless	III	1.51	68.0
Loopful	I, II, III, IV, V, and VI	0.83	62.5

* One unit of activity is defined as the enzyme activity that produced one μmol of GlcNAc per minute at 37°C. Melting temperature (T_m) was measured using differential scanning fluorimetry (DSF) followed by the procedures as described in the Methods.

Acknowledgments

Financial support by Okinawa Institute of Science and Technology (OIST) is gratefully acknowledged. We thank Nobuhiko Tokuriki for the discussion on the project. We thank Toki Taira, and Ben Clifton and Erika Fukuhara for their insightful comments on this manuscript. We are grateful for the help provided by the scientific computing, the Data Analysis, and the Instrumental sections at OIST. We also thank Paolo Barzaghi from the Imaging Section of OIST for the help with microscope experiments and Stefano Pascarelli for assistance with molecular dynamics simulations. The synchrotron radiation experiments were performed at BL45XU and BL41XU of SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (Proposal No. 2022A2769)

Author contributions

D.K. and P.L. designed the project. D.K. performed phylogenetic analysis, all the biochemical experiments, biophysical characterization of all the proteins and mutants, fluorescent imaging microscope experiments, and molecular dynamics simulations. A.S. performed rigidity-transmission allostery computation and the related data analysis. D.K. and P.L. wrote the manuscript with the input from A.S. This project was supervised by P.L.

References

1. Arnold, F. H. Innovation by evolution: bringing new chemistry to life (nobel lecture). *Angew. Chem. Int. Ed* **58**, 14420–14426 (2019).

2. Tawfik, D. S. Biochemistry. Loop grafting and the origins of enzyme species. *Science* **311**, 475–476 (2006).

3. Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D. & Aravind, L. Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J. Mol. Biol.* **361**, 1003–1034 (2006).

4. Akiva, E., Copp, J. N., Tokuriki, N. & Babbitt, P. C. Evolutionary and molecular foundations of multiple contemporary functions of the nitroreductase superfamily. *Proc Natl Acad Sci USA* **114**, E9549–E9558 (2017).

5. Monzingo, A. F., Marcotte, E. M., Hart, P. J. & Robertus, J. D. Chitinases, chitosanases, and lysozymes can be divided into procaryotic and eucaryotic families sharing a conserved core. *Nat. Struct. Biol.* **3**, 133–140 (1996).

6. Park, H.-S. *et al.* Design and evolution of new catalytic activity with an existing protein scaffold. *Science* **311**, 535–538 (2006).

7. Afriat-Jurnou, L., Jackson, C. J. & Tawfik, D. S. Reconstructing a missing link in the evolution of a recently diverged phosphotriesterase by active-site loop remodeling. *Biochemistry* **51**, 6047–6055 (2012).

8. Miton, C. M. & Tokuriki, N. Insertions and deletions (indels): A missing piece of the protein engineering jigsaw. *Biochemistry* (2022) doi:10.1021/acs.biochem.2c00188.

9. Brejc, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269–276 (2001).

10. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002).

11. Fushinobu, S., Nishimasu, H., Hattori, D., Song, H.-J. & Wakagi, T. Structural basis for the bifunctionality of fructose-1,6-bisphosphate aldolase/phosphatase. *Nature* **478**, 538–541 (2011).

12. Hoque, M. A. *et al.* Stepwise loop insertion strategy for active site remodeling to generate novel enzyme functions. *ACS Chem. Biol.* **12**, 1188–1193 (2017).

13. Dodani, S. C. *et al.* Discovery of a regioselectivity switch in nitrating P450s guided by molecular dynamics simulations and Markov models. *Nat. Chem.* **8**, 419–425 (2016).

14. Schlumbaum, A., Mauch, F., Vögeli, U. & Boller, T. Plant chitinases are potent inhibitors of fungal growth. *Nature* **324**, 365–367 (1986).

15. Taira, T. *et al.* Cloning and characterization of a small family 19 chitinase from moss (Bryum coronatum). *Glycobiology* **21**, 644–654 (2011).

16. Sljoka, A. Probing Allosteric Mechanism with Long-Range Rigidity Transmission Across Protein Networks. *Methods Mol. Biol.* **2253**, 61–75 (2021).

17. Whiteley, W. Counting out to the flexibility of molecules. *Phys. Biol.* **2**, S116-26 (2005).

18. Sljoka, A. Structural and functional analysis of proteins using rigidity theory. in *Sublinear computation paradigm: algorithmic revolution in the big data era* (eds. Katoh, N. et al.) 337–367 (Springer Singapore, 2022). doi:10.1007/978-981-16-4095-7_14.

19. Huang, S. K. *et al.* Delineating the conformational landscape of the adenosine A2A receptor during G protein coupling. *Cell* **184**, 1884-1894.e14 (2021).

20. Mehrabi, P. *et al.* Substrate-Based Allosteric Regulation of a Homodimeric Enzyme. *J. Am. Chem. Soc.* **141**, 11540–11556 (2019).

21. Baksh, K. A., Augustine, J., Sljoka, A., Prosser, R. S. & Zamble, D. B. Mechanistic insights into the nickel-dependent allosteric response of the Helicobacter pylori NikR transcription factor. *J. Biol. Chem.* 102785 (2022) doi:10.1016/j.jbc.2022.102785.

22. Ye, L. *et al.* Mechanistic insights into allosteric regulation of the A2A adenosine G protein-coupled receptor by physiological cations. *Nat. Commun.* **9**, 1372 (2018).

23. Jacobs, D. J., Rader, A. J., Kuhn, L. A. & Thorpe, M. F. Protein flexibility predictions using graph theory. *Proteins* **44**, 150–165 (2001).

24. Chothia, C., Gough, J., Vogel, C. & Teichmann, S. A. Evolution of the protein repertoire. *Science* **300**, 1701–1703 (2003).

25. Rockah-Shmuel, L. *et al.* Correlated occurrence and bypass of frame-shifting insertion-deletions (InDels) to give functional proteins. *PLoS Genet.* **9**, e1003882 (2013).

26. Emond, S. *et al.* Accessing unexplored regions of sequence space in directed enzyme evolution via insertion/deletion mutagenesis. *Nat. Commun.* **11**, 3469 (2020).

27. Schenkmayerova, A. *et al.* Engineering the protein dynamics of an ancestral luciferase. *Nat. Commun.* **12**, 3616 (2021).

28. Ross, C. M., Foley, G., Boden, M. & Gillam, E. M. J. Using the Evolutionary History of Proteins to Engineer Insertion-Deletion Mutants from Robust, Ancestral Templates Using Graphical Representation of Ancestral Sequence Predictions (GRASP). *Methods Mol. Biol.* **2397**, 85–110 (2022).

29. Macdonald, C. B. *et al.* Deep insertion, deletion, and missense mutation libraries for exploring protein variation in evolution, disease, and biology. *BioRxiv* (2022) doi:10.1101/2022.07.26.501589.

30. Campbell, E. *et al.* The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* **12**, 944–950 (2016).

31. Tóth-Petróczy, A. & Tawfik, D. S. Protein insertions and deletions enabled by neutral roaming in sequence space. *Mol. Biol. Evol.* **30**, 761–771 (2013).

32. Tokuriki, N. *et al.* Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat. Commun.* **3**, 1257 (2012).

33. Noda-Garcia, L. & Tawfik, D. S. Enzyme evolution in natural products biosynthesis: target- or diversity-oriented? *Curr. Opin. Chem. Biol.* **59**, 147–154 (2020).

34. Kaltenbach, M., Emond, S., Hollfelder, F. & Tokuriki, N. Functional Trade-Offs in Promiscuous Enzymes Cannot Be Explained by Intrinsic Mutational Robustness of the Native Activity. *PLoS Genet.* **12**, e1006305 (2016).

35. Arantes, V. & Saddler, J. N. Cellulose accessibility limits the effectiveness of minimum cellulase loading on the efficient hydrolysis of pretreated lignocellulosic substrates. *Biotechnol. Biofuels* **4**, 3 (2011).

36. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

37. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

38. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).

39. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

40. Nozaki, S. & Niki, H. Exonuclease III (XthA) Enforces In Vivo DNA Cloning of Escherichia coli To Create Cohesive Ends. *J. Bacteriol.* **201**, (2019).

41. Imoto, T. & Yagishita, K. A simple activity measurement of lysozyme. *Agric. Biol. Chem.* **35**, 1154–1156 (1971).

42. Hirata, K. *et al.* ZOO: an automatic data-collection system for high-throughput structure analysis in protein microcrystallography. *Acta Crystallogr. D Struct. Biol.* **75**, 138–150 (2019).

43. Yamashita, K., Hirata, K. & Yamamoto, M. KAMO: towards automated data processing for microcrystals. *Acta Crystallogr. D Struct. Biol.* **74**, 441–449 (2018).

44. Kabsch, W. *XDS*. in *International Tables for Crystallography: Crystallography of biological macromolecules* (eds. Arnold, E., Himmel, D. M. & Rossmann, M. G.) vol. F 304–310 (International Union of Crystallography, 2012).

45. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).

46. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).

47. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).

48. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

49. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

50. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

51. Páll, S. *et al.* Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* **153**, 134110 (2020).

52. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).

53. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

54. Parrinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182 (1981).

55. Jacobs, D. J. & Hendrickson, B. An Algorithm for Two-Dimensional Rigidity Percolation: The Pebble Game. *J. Comput. Phys.* **137**, 346–365 (1997).