Title: Evolution of oligomeric state through allosteric pathways that mimic ligand binding

Authors: Tina Perica[1,2,§], Yasushi Kondo[2], Sandhya P. Tiwari[3], Stephen H. McLaughlin[2], Katherine R. Kemplen[4], Xiuwei Zhang[1], Annette Steward[4], Nathalie Reuter[3], Jane Clarke[4] and Sarah A. Teichmann[1]*

Affiliations:

[1] European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.

[2] MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK.

[3] Department of Molecular Biology, University of Bergen and Computational Biology Unit, Department of Informatics, University of Bergen, PO Box. 7803, N-5020 Bergen, Norway.

[4] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK.

[§] Current address: Department of Bioengineering and Therapeutic Sciences, California Institute for Quantitative Biosciences, University of California, San Francisco, 1700 4th St., San Francisco, CA 94143, USA

*Correspondence to: saraht@ebi.ac.uk

Abstract: Evolution and design of protein complexes is almost always viewed through the lens of amino acid mutations at protein interfaces. We showed previously that residues not involved in the physical interaction between proteins make important contributions to oligomerisation by acting indirectly or allosterically. Here, we sought to investigate the mechanism by which allosteric mutations act using the example of the PyrR family of pyrimidine operon attenuators. In this family, a perfectly sequence-conserved helix that forms a tetrameric interface is exposed as solvent-accessible surface in dimeric orthologues. This means that mutations must be acting from a distance to destabilize the interface. We identified eleven key mutations controlling oligomeric state, all distant from the interfaces and outside ligand-binding pockets. Finally, we show that the key mutations introduce conformational changes equivalent to the conformational shift between the free versus the nucleotide-bound conformations of the proteins.

One Sentence Summary: This work probes the mechanism of indirect, allosteric mutations that employ the intrinsic dynamics of the protein involved in allosteric regulation by small molecules.

Main Text:

Introduction

Proteins diverge during the course of evolution and experience a continuous trade-off between selection for function and stability (*1*). Gould and Lewontin described how organisms adapt to different competing demands, while at the same time accumulating traits that occur either due to drift or correlations with selected features (*2*). This view can also be applied to proteins, where mutations of individual residues interact and determine fitness, similar to mutations in genes at the level of organisms (*3*). Selection is then determined by conditions, both internal (interactions with other macromolecules in the cell), and external (environmental variables, e.g. temperature or pH). Furthermore, due to the difference in the sizes of sequence and structure space, proteins can accumulate destabilizing mutations, as long as they remain stable enough at given conditions (*4*).

In previous work, we showed that mutations outside protein interfaces are as important for the evolution of quaternary structure/oligomeric state as mutations directly within interfaces (*5*). This raised the following question: by what mechanism do mutations outside interfaces affect their formation? The most likely hypothesis is that these mutations act by changing either protein conformation or conformational dynamics, analogous to the ways in which allosteric ligands introduce conformational change. Thus we referred to the indirect mutations as *allosteric mutations*.

Furthermore, the conformational dynamics of proteins enable functional features such as ligand binding, and also contribute to evolutionary plasticity, *i.e.* "evolvability". Protein dynamics are essential for the functions of many proteins (*6*), and are more conserved at the superfamily level

than sequence (*7*). Selection favours mutations of side chain interactions that promote acquisition of the folded state. In the same way, selection is stronger on functionally relevant conformations of the entire protein structural ensemble (*8*). Importantly, the conformation under strongest constraint is not the one with the lowest free energy, but rather the one most similar to the functional, often ligand-bound, state.

The protein family we study here is a group of pyrimidine attenuator regulatory proteins, PyrR, present in the *Bacillaceae* family as well as in some other bacterial species (*9*). The PyrR family shows clear evidence of mutations acting allosterically with respect to the protein interface. The change from homodimeric to homotetrameric family members is unmistakably brought about by allosteric mutations: homologues with different oligomeric states share a helix whose surface is 100% conserved in sequence. This helix forms the tetrameric interface in the homotetrameric family members, but is solvent-exposed in the dimeric family members.

Bacillaceae live at a wide range of temperatures, to which the PyrR proteins have adapted. At the same time PyrR is constrained by the need to conserve its dsRNA-binding ability and allosteric regulation by nucleotides. PyrR binds to a stem-loop structure in the nascent mRNA of the *pyr* operon, which induces formation of the termination loop and attenuates transcription. UMP and GMP allosterically regulate binding of PyrR to RNA, reflecting the ratio between purines and pyrimidines in the cell (*10, 11*) (Figure 1). As the name suggests, excess pyrimidines as reflected in UMP binding attenuate further transcription of the pyr operon.

In this work we use ancestral sequence reconstruction to infer the allosteric mutations that changed the oligomeric state and thermostability in the PyrR family during the course of evolution. We identify eleven allosteric mutations that decrease thermostability in all PyrR proteins, but change the oligomeric state only in the context of inferred ancestral PyrR proteins,

and not in thermophilic PyrR. We show how these mutations affect oligomeric state indirectly, and describe this allosteric mechanism: the same internal conformational switch in PyrR proteins is toggled both by an allosteric ligand (GMP), and by a small number of mutations.

Results and discussion

Close homologues of PyrR have conserved interface amino acids but different oligomeric states.

Using size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) at room temperature and velocity analytical ultracentrifugation (AUC) at 10 °C, we observed that the PyrR oligomeric state differs between *B. caldolyticus* and *B. subtilis* homologues, and is affected by allosteric regulators such as GMP (Figure 2). *Bacillus caldolyticus* has an optimal growth temperature of 72 °C, and its PyrR (BcPyrR) elutes as one peak corresponding to a tetramer (Figure 2). Velocity AUC experiments show that the majority of BcPyrR sediments as a tetramer with a minor monomeric species and no apparent dimeric species (Figure S4). *Bacillus subtilis* has an optimal growth temperature of 25 °C, and BsPyrR elutes from the size-exclusion column as a broad peak with a range of molecular masses between monomeric, dimeric and tetrameric oligomeric state (Figure 2). In the velocity AUC experiment for BsPyrR, two species were observed to sediment, calculated to have molecular masses corresponding to that of a PyrR dimer and tetramer respectively (Figure S4). Therefore, BsPyrR exists as a dimer in equilibrium with the monomeric and tetrameric species at low micromolar concentrations, which correspond to the physiological range, as the estimated average concentration of PyrR in *B. subtilis* cells is 0.4 μM (*12*).

BcPyrR and BsPyrR have the highest sequence identity of all PyrR homologues of known 3D structure with different oligomeric states: 73% sequence identity over 180 residues, corresponding to 49 substitutions, most of which are on the solvent-exposed surface of the protein. Interestingly, the residues involved in the tetrameric (dimer-of-dimers) interface are 100% sequence-identical. These interface residues are also likely involved in RNA-binding (*13*), and hence under purifying selection (Figure S2).

A small number of allosteric mutations change the oligomeric state of PyrR

The PyrR protein is present in various *Bacillus* species with diverse optimal growth temperatures, as well as the distantly related bacteria *Mycobacterium tuberculosis* and *Thermus thermophilus,* as shown in the phylogenetic tree of this protein family (Figure 1B). In order to trace the mutations changing the oligomeric state between the thermophilic BcPyrR (red) and mesophilic BsPyrR (blue), we focused on the two internal nodes in the phylogenetic tree after the split of the BcPyrR from the last common ancestor of the *Bacillus sp*. PyrR (LCABacillusPyrR). We reconstructed the most likely ancestral sequences of the internal nodes (*14*). Please refer to the Methods for details. In analogy to the colour wheel, we named the two inferred ancestral proteins AncORANGEPyrR and AncGREENPyrR (Figure 1B)

SEC-MALS revealed AncORANGEPyrR formed a stable tetramer, and AncGREENPyrR only showed a decrease in the average molecular mass at the lowest concentration (1 μM) from which we imply a presence of a low concentration of lower oligomeric state species (Figure 2). In AUC analysis, both ancestral proteins displayed similar distributions of sedimenting species, as seen for BcPyrR (Figure S4). Therefore, we infer that evolution of the dimeric state occurred toward the terminal branches of the PyrR phylogenetic tree, between AncGREENPyrR and BsPyrR. There are twelve substitutions and three insertions/deletions (a set of fifteen mutations we refer

to as m_3) between AncGREENPyrR and BsPyrR (Figure S19). As our goal is to identify the smallest subset of allosteric mutations that clearly change the oligomeric state we excluded four of these mutations: three insertions/deletions and one (E4Q) substitution which is a revertant to the same amino acid as in the tetrameric BcPyrR. Two of the insertions/deletions were at each of the termini, and the third one was in a flexible loop. We would not expect these three changes to have a significant effect on the structure. Eleven substitutions ($11/m_3$) were different between BsPyrR and all the tetrameric PyrRs (BcPyrR, AncORANGEPyrR and AncGREENPyrR). In order to confirm their role in the shift of the oligomeric state, we inserted them into the stable PyrR tetramer AncORANGEPyrR, producing the engineered protein VIOLETPyrR (Figure 3). The eleven substitutions did indeed destabilize the AncORANGEPyrR tetramer: VIOLETPyrR has similar SEC-MALS and AUC profiles as BsPyrR (Figures 2 and S4). Thus, remarkably, these mutations, none of which are in the tetrameric interface, shift the oligomeric state through an indirect, allosteric mechanism.

Do these mutations turn any PyrR homologue into a dimer? We grafted the eleven allosteric mutations into the tetrameric BcPyrR, forming the engineered protein PURPLEPyrR. Surprisingly, PURPLEPyrR remains tetrameric, even at lowest concentration (Figures 3B and S4). This implies that these eleven allosteric mutations have an epistatic interaction with the 32 m_1 mutations that separate BcPyrR and AncORANGEPyrR. Epistasis between amino acid substitutions is known to be ubiquitous in proteins, as described in multiple recent publications (*3, 15, 16*).

In order to pinpoint the key oligomeric state-switching mutations, we further tested the effect of two non-overlapping sets of residues within the eleven allosteric mutations, represented by PLUMPyrR ($3/m_3$) and MAGENTAPyrR ($8/m_3$). We selected the three mutations in PLUMPyrR

based on the proximity of these residues to the dimeric interface, expecting them to have the largest impact on the inter-subunit geometry. Both subsets of mutations had independent effects on the oligomeric state (Figure 3), as the equilibria of both PLUMPyrR and MAGENTAPyrR were shifted towards the dimeric state at the lowest protein concentrations in SEC-MALS, with the appearance of a dimeric species in AUC (Supp. Fig 4). This implies that these two small sets of mutations contribute to the oligomeric shift in a cumulative manner.

The eleven mutations that shift PyrR oligomeric state are part of a downhill adaptation to temperature

Members of the *Bacillus* genus live in dramatically different environments, the most notable difference being ambient temperature. Hobbs *et al* (*17*) have shown that the *Bacillus* species have adapted to different temperatures multiple times in evolution. *B. subtilis* (with the homodimeric BsPyrR) lives in soil with optimal growth at 25 °C, and *B. caldolyticus* (with the homotetrameric BcPyrR) in alkaline hot springs with the optimal growth at 72 °C (*18*).

We recorded the circular dichroism (CD) spectra at temperatures from 20 to 90 °C for all of our PyrR constructs (Figure 3C). Their thermal unfolding was irreversible, and all but BsPyrR and PURPLEPyrR unfolded in a single phase. This was sufficient to estimate the thermal stability of PyrR proteins along the evolutionary tree. BcPyrR shows no variation in the CD spectra up to 70 °C, when it suddenly unfolds cooperatively. BsPyrR however, exhibits changes in helicity at temperatures as low as 35 °C, finally unfolding completely at 75 °C. We could not determine from the CD spectra which of the secondary structure changes occur at low temperatures in BsPyrR. However, plotting ellipticity for different wavelengths suggests an exchange between α helical and β sheet structure (Fig. S5).

AncORANGEPyrR thermal unfolding follows the same pattern as that of BcPyrR, with unfolding taking place at 80 °C rather than 70 °C. This stabilization of 10 °C is most probably an artefact of ancestral protein sequence reconstruction, which has been suggested to overestimate protein stability due to a bias towards more stabilizing mutations in the evolutionary substitution models (*19*). Notably, VIOLETPyrR, which differs from AncORANGEPyrR by just the eleven mutations, unfolded at a significantly lower temperature than AncORANGEPyrR, while PLUMPyrR and MAGENTAPyrR have intermediate thermostability (Figure 3C).

This raises the question as to the mechanism for the thermal destabilizing effect of the eleven mutations. They could be affecting thermal destabilisation through the switch in oligomeric state, or by changing the polarity of the protein surface. Both residue composition and oligomeric state have been suggested to play a role in protein thermostability (*20*). A higher oligomeric state is proposed to increase thermostability by burying more residue surface area. It has also been repeatedly observed that thermophilic bacteria have more charged residues and fewer polar residues compared to mesophilic organisms. This bias is especially pronounced when only surface residues are taken into account (*21*).

To deconvolute whether it is the residue propensity or the oligomeric state that plays the main role in the differences in thermostability of PyrR, we took advantage of the engineered PURPLEPyrR, which has the eleven allosteric mutations, but is still tetrameric. Although all but the eleven allosteric residues of PURPLEPyrR are the same as in the thermophilic BcPyrR, the CD measurements show that tetrameric PURPLEPyrR has significantly decreased thermostability compared to that of BcPyrR. We thus infer that it is the change in thermophilic propensity of surface residues, and not the change in the oligomeric state, that plays a major role in the change of PyrR thermostability.

In order to dissect this in detail, we bioinformatically define the residue thermophilic propensity as the log ratio of amino acid frequencies between the solvent-exposed surfaces of proteins from thermophilic and mesophilic organisms (*21*). Thus we calculated how mutations along the PyrR tree change this thermophilic propensity. As expected, the mutations increase the thermophilic propensity on the branches from AncORANGEPyrR towards BcPyrR, and decrease towards BsPyrR. Moreover, the largest decrease in thermophilic propensity occurs between AncGREENPyrR and BsPyrR, the branch that also corresponds to the switch towards the dimeric state (Figure S6).

From this, we infer that the eleven allosteric mutations are part of a more general "downhill" adaptation to lower temperatures. Thus the switch in oligomeric state of free PyrR co-occurs with the evolutionary adaptation to lower temperatures of *B. subtilis* as compared to *B. caldolyticus*. How is this dimer/tetramer switch affected by mutations that are distant from all the inter-subunit interfaces? To answer this question, we investigated the oligomeric states during allosteric regulation by ligands in this protein family.

The allosteric regulators UMP and GMP control oligomeric state

Previous *in vivo* and biochemical experiments showed that the PyrR binding to the leader RNA sequence (PyrR binding loop) of the *pyr* operon is regulated by small molecules such as UMP and GMP (*10, 11*). In summary, higher concentrations of pyrimidines increase the affinity of PyrR for the PyrR binding loop, and in turn attenuate transcription of the pyrimidine synthesis operon. Higher concentrations of purines, on the other hand, decrease the affinity for the binding loop, which in turn increases the transcription of the pyrimidine synthesis operon (*9*) (Figure 1). As allosteric regulation usually affects conformational change, we wanted to investigate how GMP and UMP influence PyrR conformation and oligomeric state.

We analysed the oligomeric state of BsPyrR and BcPyrR by SEC-MALS upon addition of allosteric ligands, and observed that both UMP and GMP stabilize the tetrameric state of PyrR (Figures 2 and S3). This is especially prominent in the case of BsPyrR, where addition of nucleotides shifts the equilibrium towards a higher oligomeric state. The RNA-bound form of PyrR, not investigated here, is likely to be dimeric, based on analytical ultracentrifugation (*11*) and mutagenesis experiments (*13*).

This means that the effect of the eleven allosteric mutations is similar to that of RNA and opposite to the nucleotide ligands, which stabilize the tetrameric state. Interestingly, the eleven allosteric mutations are also allosteric with respect to the nucleotide-binding site: each of the eleven residues is 10 Å or more away from the bound GMP molecules.

Overall, while different ligand-bound and RNA-bound forms of the protein sample both dimeric and tetrameric states, the eleven mutations shift the free protein equilibrium towards the dimeric state in this landscape of different conformations.

Both mutations and ligands shift oligomeric state by changing inter-subunit geometry.

In order to determine the structural changes that occur when the allosteric mutations switch the oligomeric state in the PyrR family, we solved four new X-ray crystal structures: AncORANGEPyrR, AncGREENPyrR, VIOLETPyrR, and BsPyrR+GMP (Table S2). We then compared these structures to those of BcPyrR and BsPyrR (*22, 23*).

In our previous work, we hypothesized that evolutionary changes in oligomeric state can arise from difference in inter-subunit geometry within a protein complex (*5*). If this were true for the PyrR family, we would expect the dimeric structures to have distinct inter-subunit geometries as compared to the tetrameric structures.

Superimposing the dimeric BsPyrR on the BcPyrR tetramer shows an 8° rotation around the dimeric interface (Figures 4A and S1). This conformation of BsPyrR is not compatible with the tetrameric oligomeric state, as the two helices that would form the dimer-of-dimers interface are pulled apart by more than 5 Å. The eleven mutations affect the same difference in conformation. Dimeric VIOLETPyrR, and tetrameric AncORANGEPyrR, which differ by the eleven mutations, exhibit the same relative rotations between subunits within the dimer (Figure 4). The subset of three allosteric mutations introduced into PLUMPyrR from AncORANGEPyrR leads to the same geometric change, where PLUMPyrR has a 9° inter-subunit rotation as compared to AncORANGEPyrR (Figure S8).

How does this compare to the difference between free, dimeric BsPyrR and tetrameric GMP-bound BsPyrR? Addition of GMP introduces a 10° rotation, changing the protein conformation into the one compatible with forming the dimer of dimers (Figure 4). The tetrameric GMP-bound BsPyrR structure is similar to the tetramers formed by free BcPyrR and AncORANGEPyrR (Figure S8). There is a subtle 3.6 ° subunit rotation around the dimeric interface between the GMP bound form of BsPyrR, and AncORANGEPyrR. However, their tetrameric interfaces superimpose almost perfectly, with an average atomic distance difference in the tetrameric interface helices of less than 1 Å.

In summary, homologous dimers all have a similar set of inter-subunit geometries, equally different from the inter-subunit geometries of tetramers. The tetramers exhibit limited variations in their geometries, all of which are significantly smaller than the differences between the dimers and tetramers. As the geometries of the dimers and tetramers are so clearly distinct from each other, we conclude that the eleven allosteric mutations affect oligomeric state in a manner almost identical to the allosteric ligand GMP.

How does this change in intersubunit geometry come about? This is not immediately evident by inspecting the individual monomeric subunits (Figure S9) or the dimeric interfaces (*5*) between the dimeric and tetrameric proteins, as they all superpose well. To look for the subtle structural differences that could account for the observed differences in conformation and oligomeric state, we used the residue – residue interaction network approach (*24-27*). With this approach, the protein structure is reduced to a network where each node represents a residue and each edge represents a physical interaction between two residues. This allows for an unbiased analysis of structures using graph theoretical methods, as illustrated in Figure S10.

The tetrameric AncORANGEPyrR and dimeric VIOLETPyrR contact networks differ in about 15% of their contacts, and these differences are non-uniformly distributed around the network (Figures 4B and S11). To estimate how much each residue contributes to the difference in residue-residue contacts, we determined the number of contact changes in two shells around the amino acid of interest. To maintain information on residue connectivity, which has been shown to determine residue evolvability (*28*), we used the absolute number of rewired residue contacts rather than normalizing by total number of contacts. This is because rewiring the contacts of a buried residue with a high connectivity will have a larger structural impact than rewiring a residue with lower connectivity.

Three out of the eleven allosteric mutations, L68I, K84D, and A118G, exhibit dramatic rewiring of contacts (Figure 4B). Specifically, when comparing AncORANGEPyrR with the dimeric structures (BsPyrR, VIOLETPyrR and PLUMPyrR), these residues rewire between one and two standard deviations more contacts than the average buried PyrR residue (Figure S12). L68 and A118 are completely buried in the protein interior, while K84 is in the more flexible part of the protein, changing its accessible surface area between different conformations.

Moreover, L68 and A118 are the two residues at the centre of the largest rewiring events in the transition from the dimeric BsPyrR to the tetrameric BsPyrR+GMP. This means that the structural changes due to the L68I and A118G mutations "mimic" the key residue rewiring events that occur upon GMP binding. Thus the evolutionary mutations and the allosteric ligand GMP share a common mechanism for achieving an identical inter-subunit rotation, leading to the same shift in oligomeric state.

The stability difference between PyrR tetramers is coupled to changes in the dynamics of dimeric units.

Above, we observed that small differences, either *via* mutation or ligand binding, affect the wiring of residue contact networks. This suggests a small energy difference between the two main conformations of PyrR. Thus we might expect both these conformations to be sampled by the vibrational normal modes that describe the intrinsic dynamics of dimeric PyrR.

We compared the similarity of the intrinsic dynamics of different PyrR dimers (both the dimeric PyrRs and the halves of tetrameric PyrRs), by comparing their flexibility calculated using elastic network modeling (ENM) (*29*). ENM provides a distribution of fluctuations around the equilibrium conformation for each structure, and the overlap of these distributions between different structures can be described by the Bhattacharyya coefficient (BC). We have previously shown that the BC ranges between 0.85 and 1 for members of the same protein family (*30*). Accordingly, the differences between all PyrR proteins are in this range (between 0.83 and 0.98).

Furthermore, it is clear that the pattern of flexibility is more similar amongst the three dimers (BsPyrR, VIOLETPyrR, PLUMPyrR) than the three tetramers (AncORANGEPyrR, PURPLEPyrR+UMP, and BsPyrR+GMP) (Figure 5A). The BsPyrR dimer and BsPyrR+GMP tetramer had a similarity score of 0.87, while BsPyrR and other dimers (VIOLETPyrR and

PLUMPyrR) had higher pairwise similarity scores reflecting more similar dynamics. Importantly, this illustrates that similarities in intrinsic dynamics amongst the PyrR proteins are not a simple function of sequence identity, given the fact that BsPyrR+GMP and BsPyrR have the same sequence and VIOLETPyrR is closer in sequence to AncORANGEPyrR than it is to BsPyrR. The clustering based on the BC score seen in Figure 5A matches the clustering based on their RMSD (Figure S14). While BC and RMSD compare different properties of structures, here the structures with the highest BC score also have lowest RMSD, which confirms that the structural changes are encoded in the intrinsic dynamics of these structures.

It has been repeatedly shown that the conformational difference, either between functional conformations or even homologues, can be sampled by a combination of a few low-frequency normal modes of the protein (*31-35*). We found that a few lowest frequency modes of PyrR proteins describe the transition between a tetramer and a dimer, both in the case of the transition being induced by allosteric ligands and by allosteric mutations.

In particular, the second lowest frequency mode of the BsPyrR+GMP protein also captured 44% of the transition from this tetramer to the dimeric BsPyrR structure. This mode is very similar to the three lowest frequency modes of tetrameric AncORANGEPyrR. The second lowest frequency mode of this tetramer also contributes most to the conformational change between AncORANGEPyrR and the dimeric VIOLETPyrR, which differ by just the eleven mutations (Figure S15C). For both tetramers, the transition to the dimer occurs *via* the low frequency normal modes that describe the same type of motions in the structure (Videos S3 – S5). This mode corresponds to the overall subunit rotation (and translation) required to go from one state to the other as described in Figure 4A.

The difference between correlations in residue motions between a dimer and a tetramer were particularly clear when comparing the correlations of dimeric and tetrameric interface residues between AncORANGEPyrR and VIOLETPyrR. In AncORANGEPyrR, the residues from the tetrameric interface exhibit correlations across the subunit and between the two tetrameric interface helices of the two subunits of a dimer, while in VIOLETPyrR the majority of the specific correlations are located close to the dimeric interface (Figures 5B and S16). Furthermore, we observed that the residues corresponding to the three out of the eleven allosteric mutations (K62P, L68I, and A118G) are within the largest regions that undergo a collective change in intrinsic dynamics from one oligomeric state to the other. The residue corresponding to the V8I mutation also experiences a notable gain in correlation in both tetramers, related to its proximity to the tetrameric interface and the overall difference seen in that region (Figure S16 C). (Details of the statistical analysis of the correlated dynamics are described in the Supplementary Material and Figure S17.)

It is important to emphasize that the dynamics was calculated for the dimeric halves of the tetrameric PyrR structures. This means that the observed differences do not stem from the additional residue contacts in the tetrameric (dimer-of-dimers) interface, but are solely due to the conformational differences between the dimer and the equivalent half of the tetramer. Thus the conformational differences between different PyrR proteins are encoded in their intrinsic dynamics and, remarkably, relatively small effects, such as ligand binding or a small number of strategic mutations, can utilise these dynamics to toggle an intrinsic conformational switch and change the quaternary structure.

Conclusions

Reconstituting the PyrR sequences and analysing their biophysical properties enables us to recapitulate the evolutionary history of the family (Figure 1B). In one part of the phylogenetic tree, PyrR has adapted to remain stable and functional at extremely high temperatures (BcPyrR), while in the other part (after the AncGREENPyrR node) the organisms have adapted to life at lower temperatures (25 °C). It is known that many proteins maintain marginal stability, and one would thus expect the protein stability to reflect the differences in environmental temperatures. This can be explained by the simple fact that the selection pressure for increased stability is relaxed at mesophilic temperatures, meaning that proteins can accumulate destabilizing mutations until they reach marginal stability (*4*).

However, high stability can come at the expense of increased conformational rigidity, and a protein adapted to be stable at high temperatures may not be flexible enough to perform its function at a lower temperature (*36*). Whether it was adaptational or simply drift, in the case of PyrR, "downhill" mutations lowered thermostability, while at the same time selection for maintaining the RNA-binding site and allosteric regulation acted continuously throughout evolution. Interestingly, the accumulated "downhill" mutations caused small and cumulative changes sufficient to switch oligomeric state in the absence of mutations in the actual tetrameric interface. This change in the stability of the tetramer may have been an evolutionary by-product, demonstrating the power and importance of indirect and structurally allosteric mutations.

We have shown that the change in oligomeric state occurred through an interplay of mutations impacting residue contact networks, inter-subunit geometry, intrinsic dynamics and thermostability. At the same time, for six out of the eleven mutations, we were able to estimate their relative contributions to each of these properties (Figure 6). Interestingly, the K84D mutation, which is in the part of the structure that seems to be disordered in some of the

conformations, is predicted to affect all three properties simultaneously. G172Q and K181N are mutations in surface residues, predicted to significantly contribute to the change in thermostability. L68I and A118G are mutations in buried residues at the centre of a large residue-residue rewiring event in the transition from dimer to tetramer, both in evolution and ligand binding. Both residues are also part of a region of the protein with highly correlated dynamics. K62P is a mutation in a surface residue, weakly connected to the rest of the structure, but part of a region with highly correlated motions where a change has a significant effect on protein dynamics.

Here we showed compellingly how mutations in residues outside the interface can introduce rearrangements that have a knock-on effect on the interface itself. We hope that the importance of mechanisms of allosteric mutations will become increasingly clear with the advancement of methods that accurately predict effects of mutations, as well as methods for engineering proteins with multiple functional conformations.

Methods

Ancestral sequence reconstruction

To reconstruct the ancestral PyrR sequences between the dimeric BsPyrR and the tetrameric BcPyrR we retrieved all the PyrR protein sequences from UniProtKB, including the sequences of two outliers: PyrR from *M. tuberculosis* and *T. thermophilus*. We used MUSCLE (*37*) to calculate a multiple sequence alignment of the PyrR proteins (Figure S18). We performed Bayesian inference with MrBayes version 3.1 (*38*). The evolutionary tree topology, branch lengths and the sequences of ancestral nodes were calculated from a PyrR protein alignment by using an estimated fixed-rate evolutionary model. The gaps in the ancestral sequences were determined using the F81-like model for binary data implemented in MrBayes (*39*). Please refer to the Supplementary Materials for more details.

Oligomeric state analysis by SEC-MALS

We resolved the protein samples on a Superdex S-200 10/300 analytical gel filtration column (GE Healthcare), pre-equilibrated with 50 mM Tris pH 7.5, 150 mM NaCl, and 1 mM DTT, at 0.5 ml/min. We performed the measurements using an online Dawn Heleos II 18 angle light scattering instrument (Wyatt Technologies Corp.) coupled to an Optilab rEX online refractive index detector (Wyatt Technologies Corp.) in a standard SEC-MALS format. We used the ASTRA v5.3.4.20 software (Wyatt Technologies Corp.) to determine the absolute molecular mass from the intercept of the Debye plot using Zimm's model (*40*) and analysed the light scattering and differential refractive index. We determined the protein concentration from the excess differential refractive index based on *dn/dc* of 0.186 mg/ml. In order to determine the

inter-detector delay volumes, band broadening constants and the detector intensity normalization constants for the instrument, we used BSA as a standard prior to sample measurement.

X- ray crystallography

All crystallisation trials were performed with 15-20 mg/ml of protein, and the sample buffer was supplemented with 10 mM $MgCl_2$. AncPURPLE crystalised with 1.2 time excess UMP and BsPyrR with 2 time excess GMP. AncGREENPyrR and BsPyrR were additionally supplemented with 400 mM $(NH_4)_2SO_4$ in order to obtain crystals. We set up 100 nl protein drop crystallisation trials with the in-house LMB screen (*41*). We collected the X-ray diffraction data at the Diamond Synchotron (Oxford, UK). The data were processed in the CCP4 suite (*42*). Please refer to Supplementary Material for details on crystallisation conditions and data processing. All the structures were solved using molecular replacement with Phaser (*43*), rebuilt with Coot (*44*), and refined with Refmac5 (*45*).

Structural superpositions and inter-subunit geometry comparisons

Inter-subunit geometries of all the PyrR structures (Figures 4 and S8) were compared as described previously in (*5*) and illustrated in Figure S1. We superimposed individual subunits using a *sievefit* approach, described by Arthur Lesk, and used notably in (*46*), and implemented more recently in the Bio3D package for R (*47*). With *sievefit*, subunits are structurally aligned using only residues that are superposable with an RMSD below 0.5 Å. These residues then define the structural core of the subunit with its corresponding centre of mass. We first *sievefit* only subunit A of the complex, and then re-*sievefit* the B subunit, noting how much the centre of mass needs to deviate from its original position (as an angle of rotation and vector of translation).

Normal mode analysis

We studied the intrinsic dynamics of all the PyrR proteins for which we had high quality structures (meaning solved to a high resolution and not having more than a few residues missing from the structure). These structures were: AncORANGEPyrR, VIOLETPyrR, PLUMPyrR, PURPLEPyrR with UMP, as well as the wild type *B. subtilis* PyrR with and without GMP. We performed the calculations using the Cα atom elastic network model implemented in the Molecular Modeling ToolKit (*48*), on the dimeric units (dimers and halves of tetramers). The GMP and UMP ligands were modelled into the *B. subtilis* PyrR and the PURPLEPyrR structures by placing dummy nodes at the C4', N9, and N1 or C4', N1, and C4 positions of the bound nucleotides in both subunits, respectively.

To compare the intrinsic dynamics of these structures, we used a structural alignment obtained from MUSTANG (*49*). The Bhattacharya coefficient (BC) (*30, 50*) was used as a measure of similarity in flexibility, with a score from 0 (completely dissimilar) to 1 (identical). Furthermore, the correlation matrices were calculated from the 100 lowest frequency modes (*51*). (For more details, refer to Supplementary Material.) The conformational overlap analysis of BsPyrR and BsPyrR+GMP, as well as AncORANGEPyrR with PLUMPyrR and VIOLETPyrR to obtain the modes that contribute to the transition from the tetrameric state to the dimeric state was done according to Reuter et al. (*34*). Here, we calculated overlaps between the modes of dimeric halves of tetramers and the structural difference vectors between the dimeric half of the tetramer and the corresponding dimer.

Thermostability

We estimated the thermostability of the PyrR proteins by measuring the circular dichroism (CD) 210-260 nm spectrum of each protein over a range of temperatures (from 20 to 90 °C). We heated the proteins gradually and continuously (0.2 °C per minute) and collected the spectrum

every 5 °C. The proteins were measured at an approximate concentration of 5 μM. All the measurements were done on a Chirascan^{TM} CD Spectrometer (AppliedPhotophysics). Mean residue ellipticity for each protein at each 5 °C temperature point was calculated as the degrees of CD corrected by exact protein concentration and the length of the protein (number of amino acids).

References

1. N. Tokuriki, D. S. Tawfik, Stability effects of mutations and protein evolvability. *Current opinion in structural biology* **19**, 596 (Oct, 2009).
2. S. J. Gould, R. C. Lewontin, The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society B: Biological Sciences* **205**, 581 (Sep 21, 1979).
3. M. Kaltenbach, N. Tokuriki, Dynamics and constraints of enzyme evolution. *Journal of experimental zoology. Part B, Molecular and developmental evolution*, (Mar 13, 2014).
4. D. M. Taverna, R. A. Goldstein, Why are proteins marginally stable? *Proteins* **46**, 105 (Feb 01, 2002).
5. T. Perica, C. Chothia, S. A. Teichmann, Evolution of oligomeric state through geometric coupling of protein interfaces. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8127 (Jun 22, 2012).
6. M. Karplus, J. Kuriyan, Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 6679 (Jun 10, 2005).
7. S. Maguid, S. Fernandez-Alberti, G. Parisi, J. Echave, Evolutionary conservation of protein backbone flexibility. *Journal of molecular evolution* **63**, 448 (Oct, 2006).
8. E. Juritz, N. Palopoli, M. S. Fornasari, S. Fernandez-Alberti, G. Parisi, Protein conformational diversity modulates sequence divergence. *Molecular biology and evolution* **30**, 79 (Feb, 2013).
9. C. L. Turnbough, R. L. Switzer, Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. *Microbiology and molecular biology reviews : MMBR* **72**, 266 (Jul 01, 2008).
10. E. R. Bonner, J. N. D'Elia, B. K. Billips, R. L. Switzer, Molecular recognition of pyr mRNA by the Bacillus subtilis attenuation regulatory protein PyrR. *Nucleic acids research* **29**, 4851 (Dec 01, 2001).
11. C. M. Jørgensen *et al.*, pyr RNA binding to the Bacillus caldolyticus PyrR attenuation protein. Characterization and regulation by uridine and guanosine nucleotides. *The FEBS journal* **275**, 655 (2008).
12. S. Maass *et al.*, Efficient, global-scale quantification of absolute protein amounts by integration of targeted mass spectrometry and two-dimensional gel-based proteomics. *Analytical chemistry* **83**, 2677 (May 01, 2011).
13. H. K. Savacool, R. L. Switzer, Characterization of the interaction of Bacillus subtilis PyrR with pyr mRNA by site-directed mutagenesis of the protein. *Journal of bacteriology* **184**, 2521 (Jun, 2002).
14. M. J. Harms, J. W. Thornton, Analyzing protein structure and function using ancestral gene reconstruction. *Current opinion in structural biology* **20**, 360 (Jul 01, 2010).
15. J. T. Bridgham, E. A. Ortlund, J. W. Thornton, An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* **461**, 515 (Sep 24, 2009).

16. L. I. Gong, M. A. Suchard, J. D. Bloom, Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).

17. J. K. Hobbs *et al.*, On the origin and evolution of thermophily: reconstruction of functional precambrian enzymes from ancestors of bacillus. *Molecular biology and evolution* **29**, 825 (Mar 01, 2012).

18. U. J. Heinen, W. Heinen, Characteristics and properties of a caldo-active bacterium producing extracellular enzymes and two related strains. *Archiv für Mikrobiologie* **82**, 1 (Jul 19, 1971).

19. P. D. Williams, D. D. Pollock, B. P. Blackburne, R. A. Goldstein, Assessing the accuracy of ancestral protein reconstruction methods. *PLoS computational biology* **2**, e69 (Jul 23, 2006).

20. M. Robinson-Rechavi, A. Alibés, A. Godzik, Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons from structural genomics of Thermotoga maritima. *Journal Of Molecular Biology* **356**, 547 (Mar 17, 2006).

21. S. Fukuchi, K. Nishikawa, Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *Journal Of Molecular Biology* **309**, 835 (Jul 15, 2001).

22. P. Chander *et al.*, Structure of the nucleotide complex of PyrR, the pyr attenuation protein from Bacillus caldolyticus, suggests dual regulation by pyrimidine and purine nucleotides. *Journal of bacteriology* **187**, 1773 (Apr 01, 2005).

23. D. R. Tomchick, R. J. Turner, R. L. Switzer, J. L. Smith, Adaptation of an enzyme to regulatory function: structure of Bacillus subtilis PyrR, a pyr RNA-binding attenuation protein and uracil phosphoribosyltransferase. *Structure (London, England : 1993)* **6**, 337 (Apr 15, 1998).

24. G. Amitai *et al.*, Network analysis of protein structures identifies functional residues. *Journal Of Molecular Biology* **344**, 1135 (Dec 03, 2004).

25. V. Soundararajan, R. Raman, S. Raguram, V. Sasisekharan, R. Sasisekharan, Atomic interaction networks in the core of protein domains and their native folds. *PLoS ONE* **5**, e9391 (2010).

26. A. J. Venkatakrishnan *et al.*, Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185 (Mar 14, 2013).

27. X. Zhang, T. Perica, S. A. Teichmann, Evolution of protein structures and interactions from the perspective of residue contact networks. *Current opinion in structural biology*, (Jul 25, 2013).

28. E. Dellus-Gur, A. Tóth-Petróczy, M. Elias, D. S. Tawfik, What Makes a Protein Fold Amenable to Functional Innovation? Fold Polarity and Stability Trade-offs. *Journal Of Molecular Biology*, (Apr 28, 2013).

29. K. Hinsen, A. J. Petrescu, S. Dellerue, M. C. Bellissent-Funel, G. R. Kneller, Harmonicity in slow protein dynamics. *Chemical Physics* **261**, 25 (2000).

30. E. Fuglebakk, J. Echave, N. Reuter, Measuring and comparing structural fluctuation patterns in large protein datasets. *Bioinformatics (Oxford, England)* **28**, 2431 (Oct 01, 2012).

31. J. Echave, Evolutionary divergence of protein structure: The linearly forced elastic network model. *Chemical Physics Letters* **457**, 413 (2008).

32. A. Leo-Macias, P. Lopez-Romero, D. Lupyan, D. Zerbino, A. R. Ortiz, An analysis of core deformations in protein superfamilies. *Biophysical journal* **88**, 1291 (Mar 01, 2005).

33. F. Raimondi, M. Orozco, F. Fanelli, Deciphering the deformation modes associated with function retention and specialization in members of the Ras superfamily. *Structure (London, England : 1993)* **18**, 402 (Apr 10, 2010).

34. N. Reuter, K. Hinsen, J.-J. Lacapère, Transconformations of the SERCA1 Ca-ATPase: a normal mode study. *Biophysical journal* **85**, 2186 (Oct, 2003).

35. F. Tama, Y. H. Sanejouand, Conformational change of proteins arising from normal mode calculations. *Protein engineering* **14**, 1 (2001).

36. P. Závodszky, J. Kardos, Svingor, G. A. Petsko, Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 7406 (Jul 23, 1998).

37. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792 (2004).

38. F. Ronquist, J. P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)* **19**, 1572 (Aug 12, 2003).

39. J. Felsenstein, Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution* **17**, 368 (1981).

40. B. H. Zimm, The Scattering of Light and the Radial Distribution Function of High Polymer Solutions. *Journal of Chemical Physics* **16**, 1093 (Dec, 1948).

41. D. Stock, O. Perisic, J. Löwe, Robotic nanolitre protein crystallisation at the MRC Laboratory of Molecular Biology. *Progress in biophysics and molecular biology* **88**, 311 (Jul, 2005).

42. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta crystallographica Section D, Biological crystallography* **67**, 235 (May, 2011).

43. A. J. McCoy *et al.*, Phaser crystallographic software. *Journal of applied crystallography* **40**, 658 (Aug 01, 2007).

44. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta crystallographica Section D, Biological crystallography* **60**, 2126 (Dec, 2004).

45. G. N. Murshudov *et al.*, REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica Section D, Biological crystallography* **67**, 355 (May, 2011).

46. M. Gerstein, C. Chothia, Analysis of protein loop closure. Two types of hinges produce one motion in lactate dehydrogenase. *Journal Of Molecular Biology* **220**, 133 (Jul 05, 1991).

47. B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, L. S. D. Caves, Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics (Oxford, England)* **22**, 2695 (Nov 01, 2006).

48. K. Hinsen, The molecular modeling toolkit: A new approach to molecular simulations. *Journal of Computational Chemistry* **21**, 79 (Feb 30, 2000).

49. A. S. Konagurthu, J. C. Whisstock, P. J. Stuckey, A. M. Lesk, MUSTANG: a multiple structural alignment algorithm. *Proteins* **64**, 559 (Aug 15, 2006).

50. E. Fuglebakk, N. Reuter, K. Hinsen, Evaluation of Protein Elastic Network Models Based on an Analysis of Collective Motions. *Journal of Chemical Theory and Computation* **9**, 5618 (Dec 10, 2013).

51. T. Ichiye, M. Karplus, Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* **11**, 205 (Nov, 1991).

Acknowledgments:

The authors wish to thank Prof. Robert L. Switzer (University of Illinois) for the cDNA of *B. subtilis* and *B. caldolyticus* PyrR. We thank Eviatar Natan and Dominika Gruszka for practical help, and Christine Vogel, Edvin Fuglebakk, and Nobuhiko Tokuriki for helpful discussions and insights. We thank Dr. Kiyoshi Nagai for generous access to his laboratory. We also thank Emmanuel Levy, Joseph Marsh, Roman Laskowski, Merridee Wouters and Boris Lenhard for feedback on the manuscript. YK was supported by Nakajima Foundation. XZ is supported by an Early Postdoc Mobility Fellowship from the Swiss National Science Foundation (SNSF, Grant Number PBELP2_143538). JC is a Senior Wellcome Trust Research Fellow (grant number 095195). ST and NR are supported by Bergen Forskningsstiftelse. This work was supported by the Medical Research Council, Lister Research Prize to SAT, and a Henry Wellcome Postdoctoral Fellowship to TP.

Coordinates and structural factors have been deposited with the Protein Data Bank (entry codes 4P80, 4P81, 4P82, 4P83, 4P84, 4P86 and 4P3K)

Figures

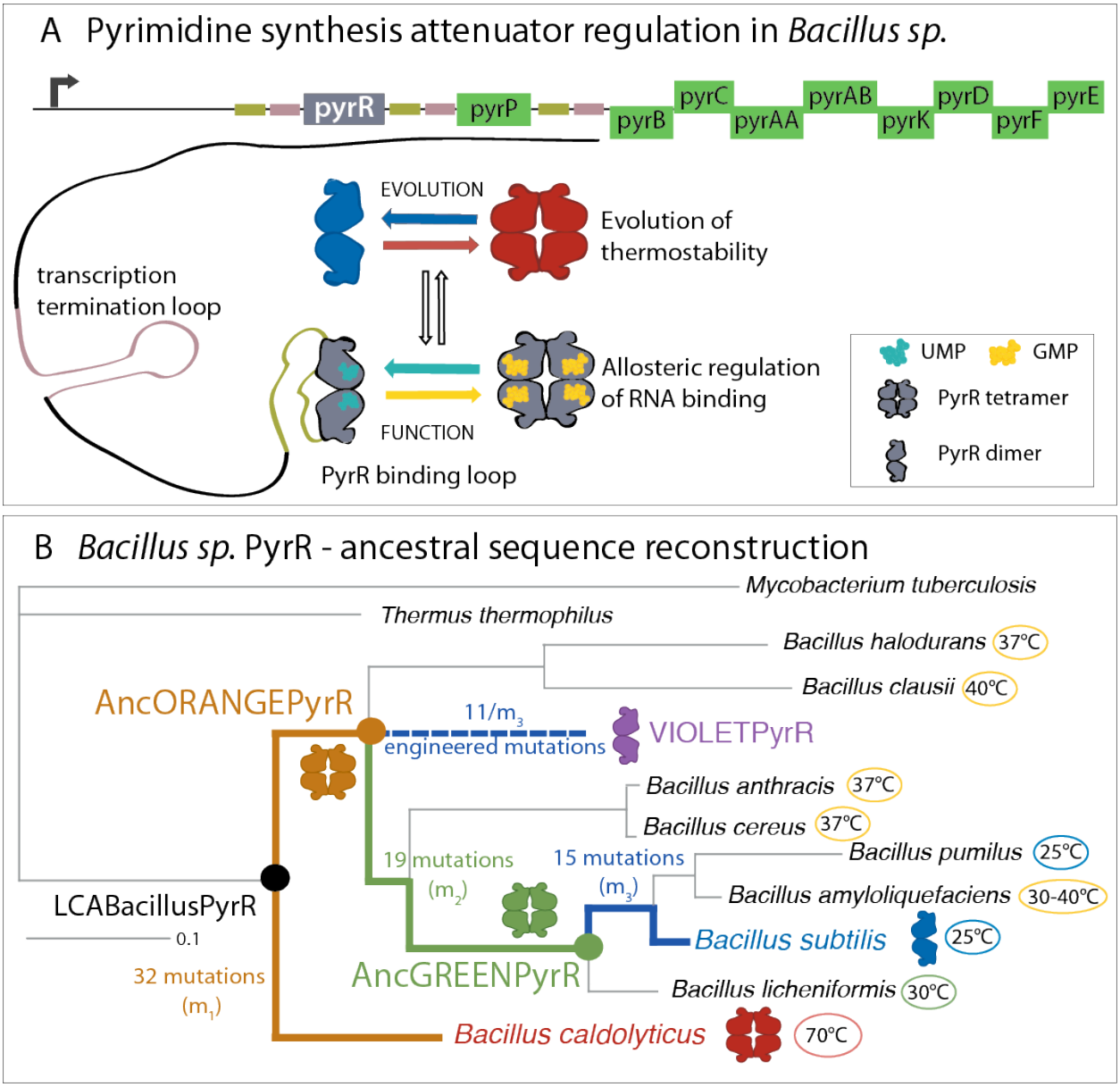


Fig. 1 (A) Schematic representation of the pyrimidine operon attenuator system in *Bacillus* sp. Attenuator protein, PyrR, binds to the PyrR binding loop as a dimer. UMP allosterically promotes the binding of RNA, while addition of GMP decreases the affinity for RNA (*11*). Different *Bacillus* species live in different environments and are adapted to different optimal growth temperatures. (B) The phylogenetic tree of *Bacillus* PyrR proteins (inferred using

Bayesian MCMC) shows the variety of optimal growth temperatures for different *Bacillus* species: *Bacillus caldolyticus*, lives at temperatures higher than 70 °C and, at room temperature, its PyrR is a homotetramer. *Bacillus subtilis* optimal growth temperature is 25 °C, and at room temperature, its PyrR is in equilibrium between a homodimer and a homotetramer (illustrated as just a dimer for simplicity). Analysis of the reconstructed ancestral sequences shows that the change from a tetramer to a dimer, occurred on the final (blue) branches of the tree, where 15 allosteric mutations (m_3) turn a tetrameric AncGREENPyrR into a dimeric BsPyrR. A subset of eleven of those allosteric mutations ($11/m_3$) also switches the oligomeric state in the context of the ancestral AncORANGEPyrR.

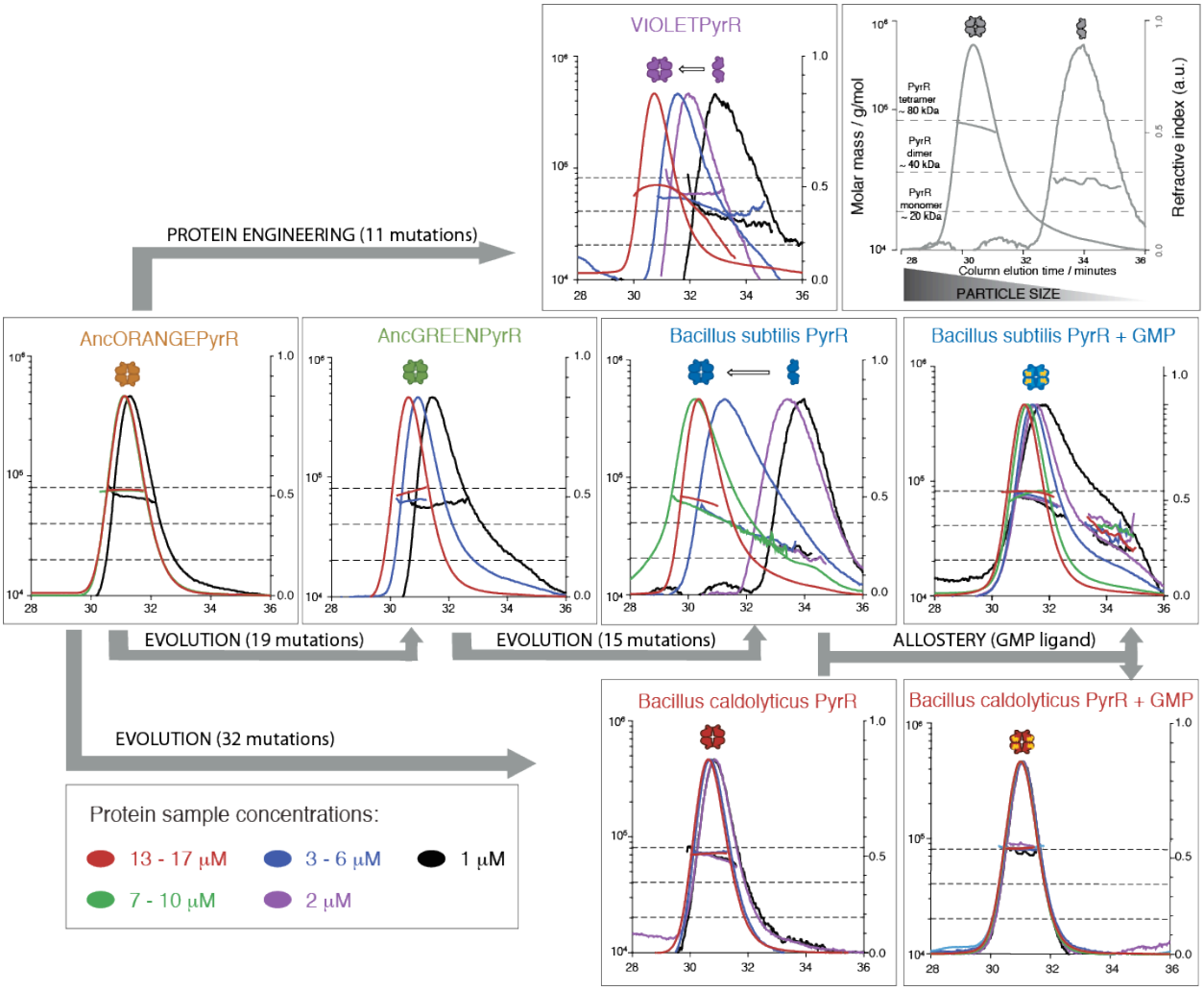


Fig. 2 Analysis of PyrR oligomeric states. Samples of AncORANGEPyrR, AncGREENPyrR, *Bacillus subtilis* PyrR (BsPyrR), BsPyrR + GMP, *Bacillus caldolyticus* PyrR (BcPyrR), BcPyrR+GMP, and VIOLETPyrR at varying concentrations were separated by size-exclusion chromatography prior to determination of the excess refractive index and multi-angle light scattering (SEC-MALS) from which the molecular masses are determined. Horizontal dashed lines represent the expected masses for monomeric, dimeric and tetrameric PyrR species, respectively.

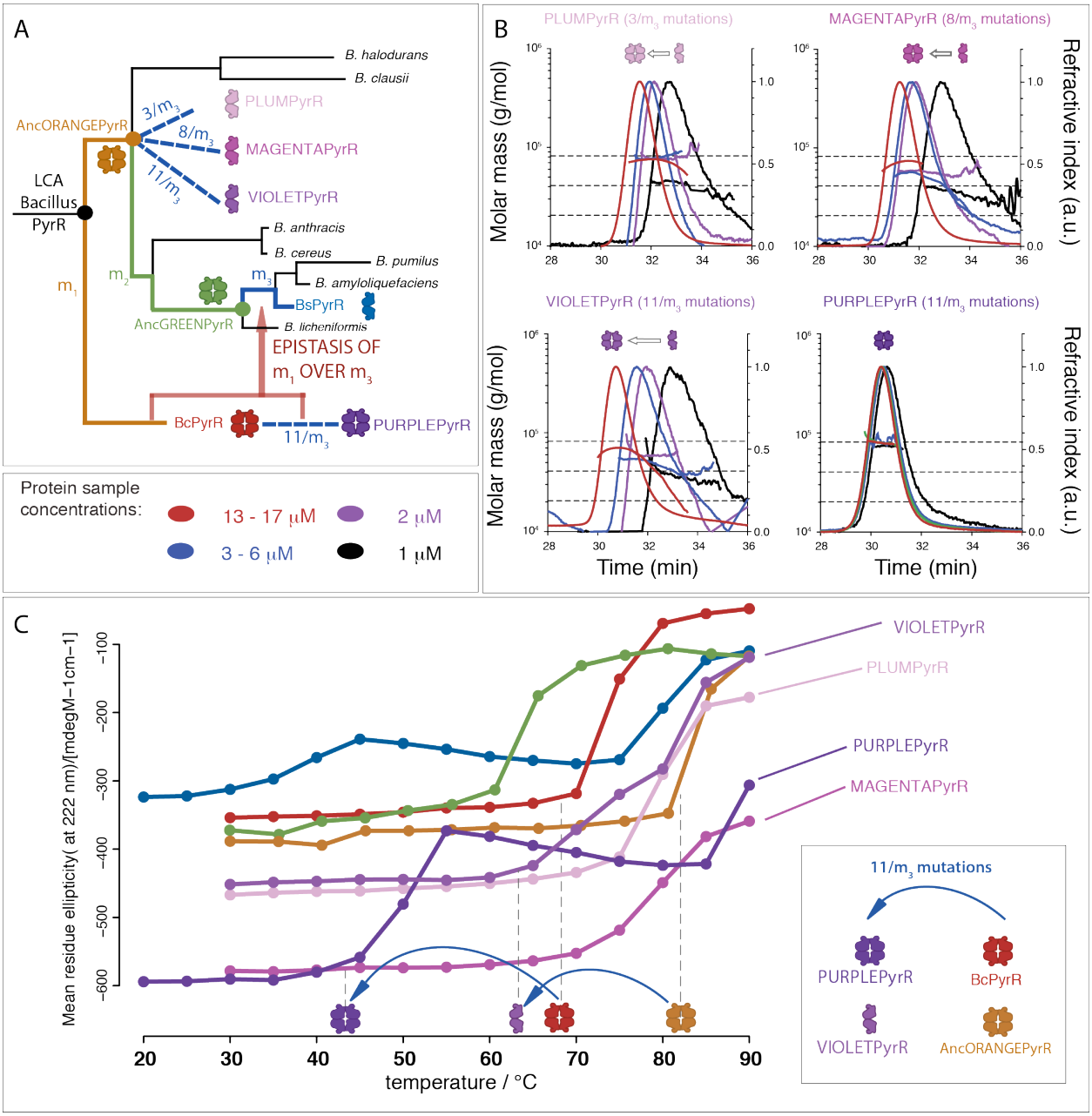


Fig. 3 Allosteric mutations affect oligomeric state and thermostability. (A) Summary of the effects of allosteric mutations on oligomeric state along the PyrR phylogenetic tree. (B) Two non-overlapping subsets of the eleven allosteric mutations (3/m_3 mutations (PLUMPyrR), and 8/m_3 mutations (MAGENTAPyrR)) are enough to overcome the threshold and cause instability of the tetramer sufficient at 1 μM protein concentrations. All eleven allosteric mutations (11/m_3)

together (VIOLETPyrR) have a larger effect on the stability of the tetramer than the $3/m_3$ and the $8/m_3$ individual subsets of mutations. The eleven allosteric mutations change oligomeric state only in the context of AncORANGEPyrR (or AncGREENyrR), but not in the context of BcPyrR. This is due to the epistasis of (a subset of) m_1 mutations over the m_3 mutations. (C) Oligomeric state and thermostability are coupled in PyrR, but it is the thermophilic propensity of residues, not the oligomeric state that determines thermostability. Thermal unfolding of PyrR homologues, inferred from circular dichroism at 222 nm at temperatures ranging from 30 to 90 °C (from 20 to 90 °C for BsPyrR and PURPLEPyrR). Loss of CD signal at 222 nm is interpreted as the loss of helicity. The circular dichroism (CD) signal is plotted as the mean residue ellipticity corrected for protein concentration.

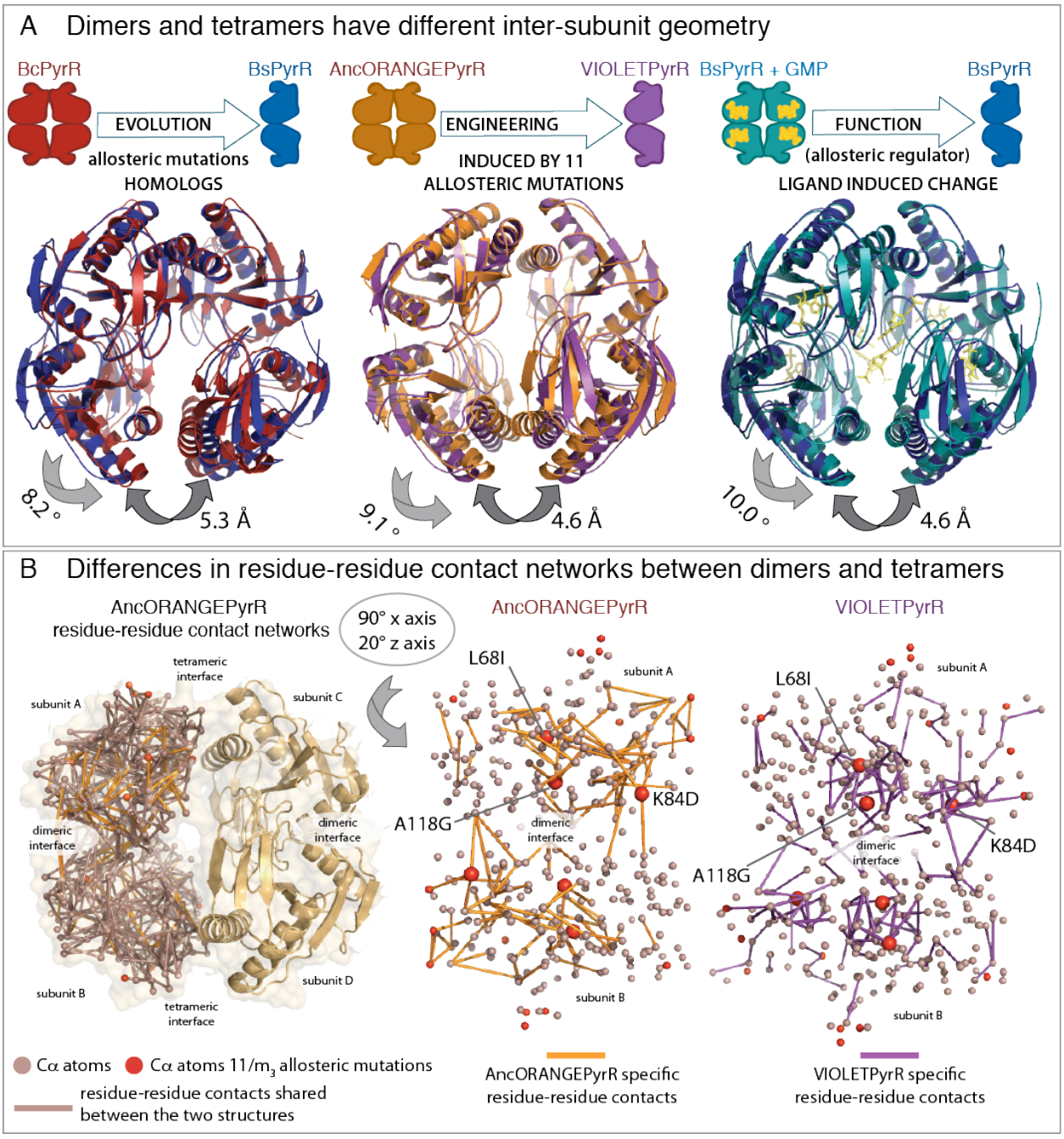


Fig. 4 (A) Change in oligomeric state through evolutionary variation, functional allostery, or recapitulated by engineering is always coupled to the same difference in inter-subunit geometry. The three superimposed pairs of structures: (i) *Bacillus subtilis* PyrR (BsPyrR, pdb:1a3c) dimers superimposed on *Bacillus caldolyticus* PyrR (BcPyrR, pdb: 1non) tetramer; (ii) VIOLETPyrR dimers superimposed on AncORANGEPyrR tetramer; (iii) *Bacillus subtilis* PyrR (BsPyrR,

pdb:1a3c) dimers superimposed on tetrameric *Bacillus subtilis* PyrR in complex with GMP. The inter-subunit geometry of free BsPyrR is incompatible with formation of the dihedral tetramer, however, GMP binding introduces a 10° inter-subunit geometric change and BsPyrR forms a tetramer. The VIOLETPyrR inter-subunit geometry is not compatible with the formation of the tetramer formed by AncORANGEpyrR, and this difference of conformation and oligomeric state is brought about by eleven allosteric mutations. (B) The tetrameric AncORANGEPyrR and dimeric VIOLETPyrR residue-residue contact networks differ by 15% of their contacts. Orientation of networks can be further explored in Supplementary Videos S1 and S2. L68I, K84D and A118G are central hubs and are involved in the majority of contact rewiring between the dimeric and tetrameric networks.

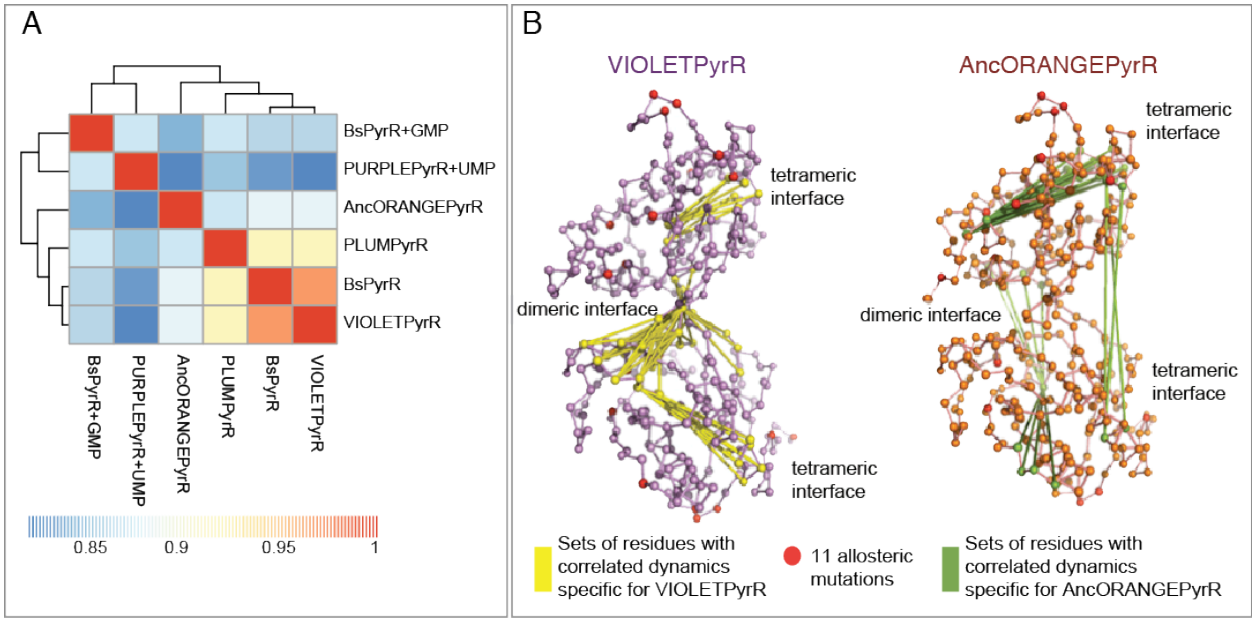


Fig. 5 PyrR intrinsic dynamics and oligomeric state. (A) All PyrR proteins have a similar intrinsic dynamics, but the three dimeric proteins are more similar than the tetramers. This difference is most pronounced when comparing the sets of dimeric and tetrameric interface residues with correlated dynamics specific for the dimeric VIOLETPyrR or the tetrameric AncORANGEPyrR. (B) Both structures are represented by only their Cα atoms, connected by green or yellow edges if at least one of the residues is involved either in the dimeric or the tetrameric interface and only if the pair of residues is moving in a concerted, correlated manner either only in the dimeric VIOLETPyrR (yellow edges) or only in the tetrameric AncORANGEPyrR (green edges). The residues corresponding to the eleven allosteric mutations ($11/m_3$) are coloured in red. The sets of residues with correlation differences shown here have a cluster size of more than three, and fall within the correlation difference threshold of 0.1. (Both threshold values were chosen for the sake of clarity; please see Figures S16 and S17 for a more exhaustive analysis of the correlation differences).

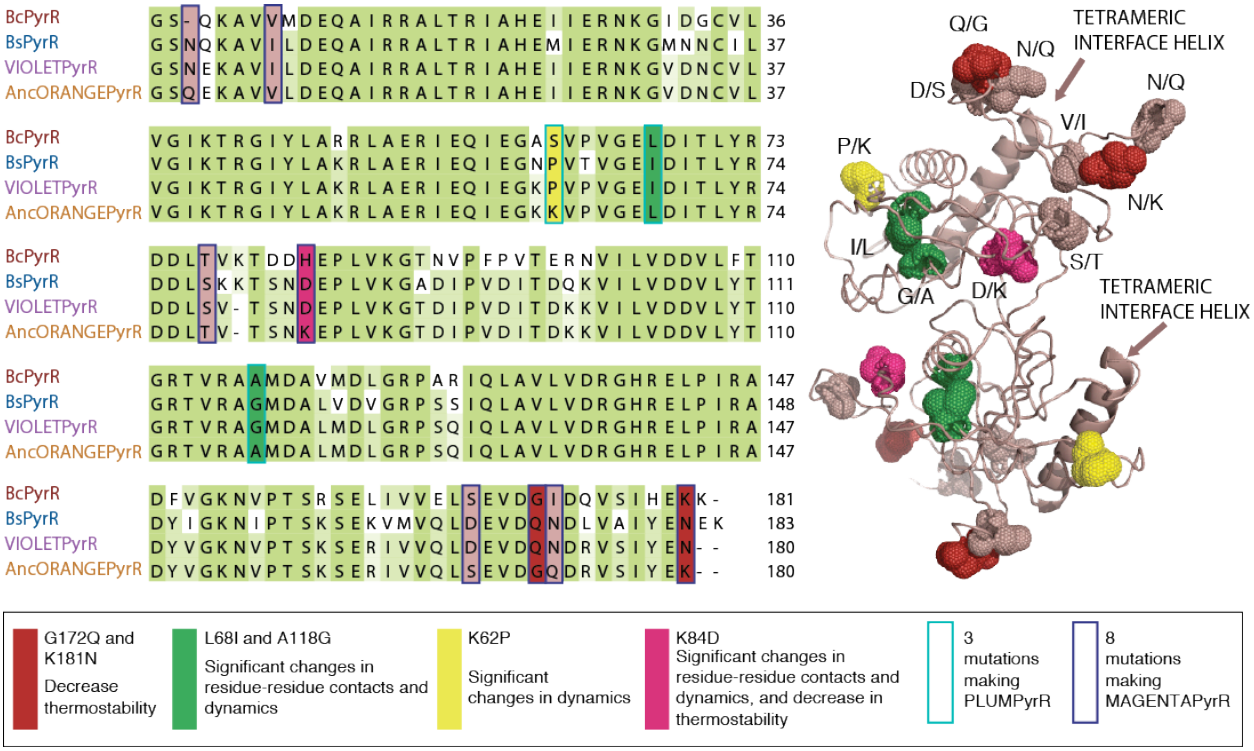


Fig. 6 Summary of mutational mechanisms. A small number of allosteric mutations are responsible for the evolutionary difference in oligomeric state, thermostability and dynamics of PyrR homologues. We show the mechanism(s) by which each mutation acts, and summarize similar mutations using the same colour.

Supplementary Materials:

Figures S1 - S18

Tables S1 and S2

Videos S1 – S4

Supplementary Materials and Methods

References 52-68

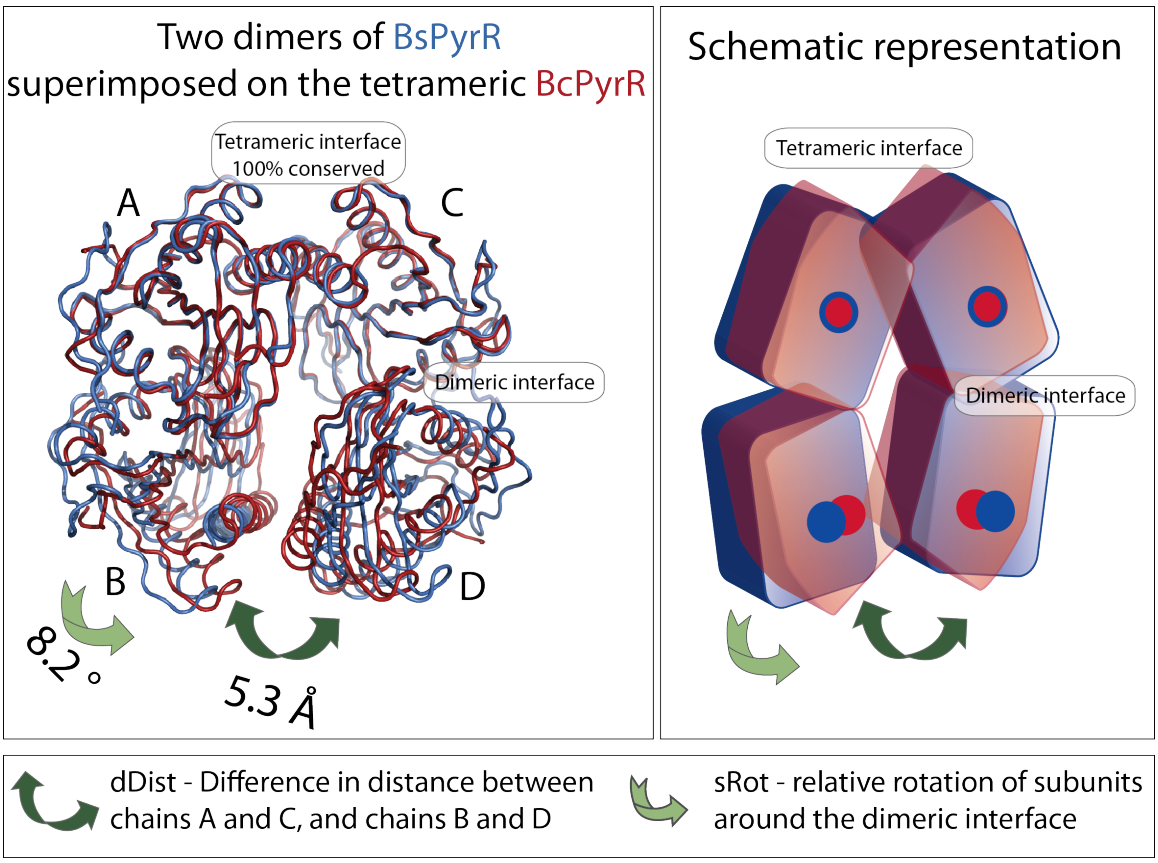


Figure S1 BcPyrR is in a conformation compatible with the formation of a homotetramer with dihedral symmetry. Conformation of BsPyrR dimer (rotated 8° around the dimeric interface) is not compatible with the BcPyrR tetramers, as the tetrameric interface helices are approximately 5 Å apart. Conformational change of relative positions of subunits (their centres of mass) around the dimeric interface (described as a rotation of approx. 8°). We compared the inter-subunit geometries of PyrR homologues using *sievefit*, an approach where subunits are structurally aligned using only residues that are superposable with an RMSD below 0.5 Å. We would first *sievefit* only subunit A of the complex, and then re-*sievefit* the B subunit, noting how much the centre of mass needs to deviate from its original position (as an angle of rotation and vector of translation).

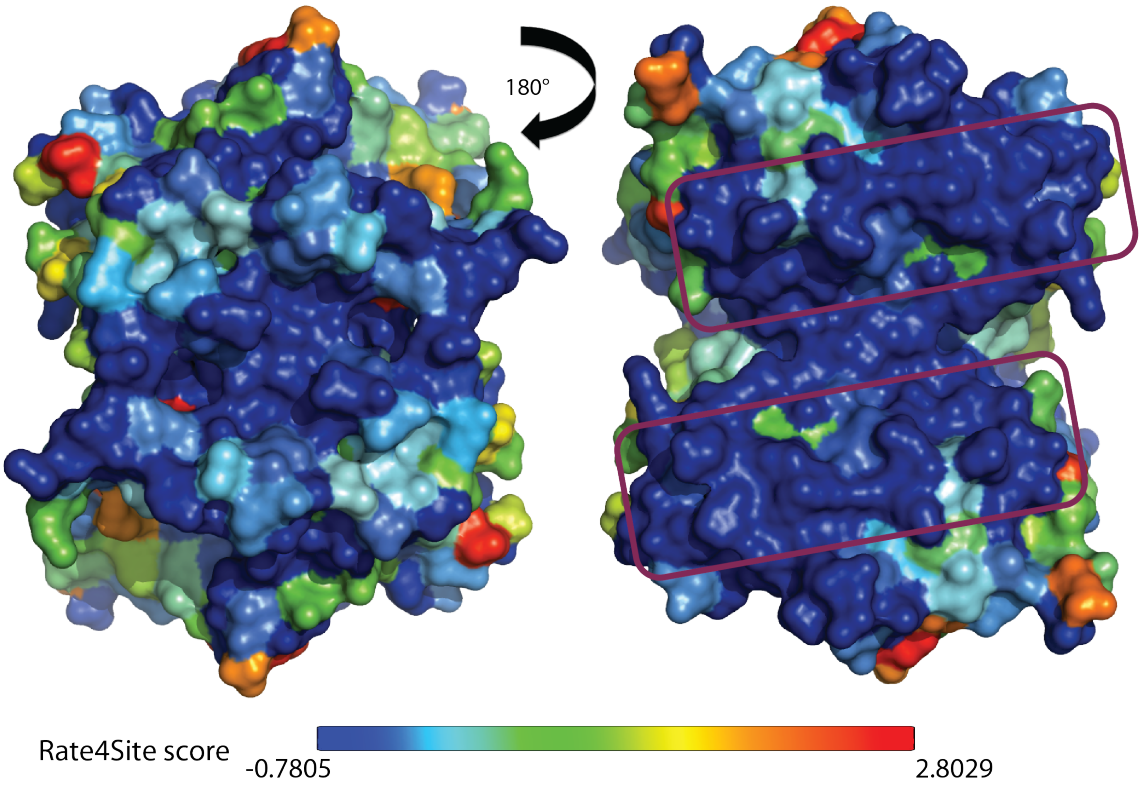


Figure S2 *Bacillus subtilis* PyrR surface residues coloured by their Rate4Site score (*52*). Residues with lower Rate4Site score are more conserved (i.e. evolve more slowly, have lower evolutionary rate). The surface on the right-hand side is more conserved and corresponds to the tetrameric interface helix (highlighted with a purple rectangular), as well as the putative RNA loop binding site.

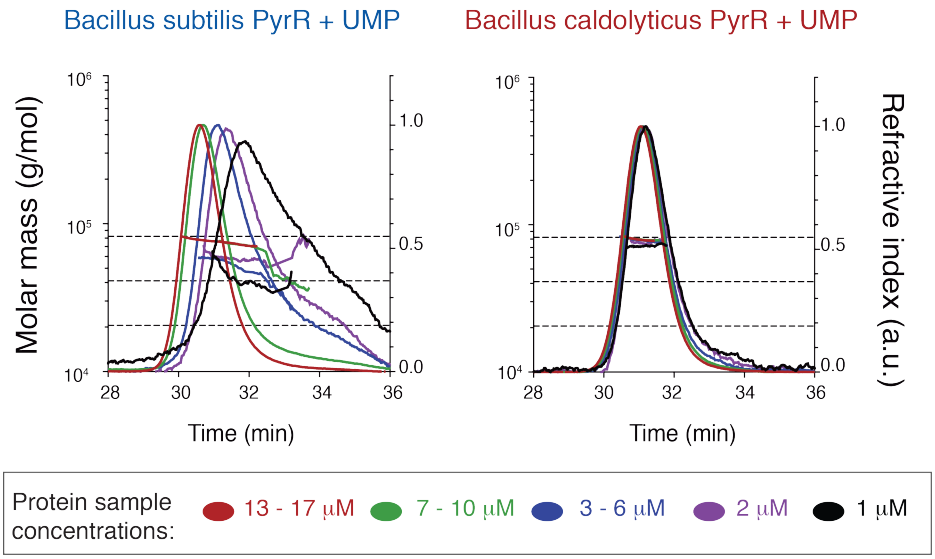


Figure S3 SEC-MALS analysis of PyrR from *B. subtilis* and *B. caldolyticus* in the presence of UMP.

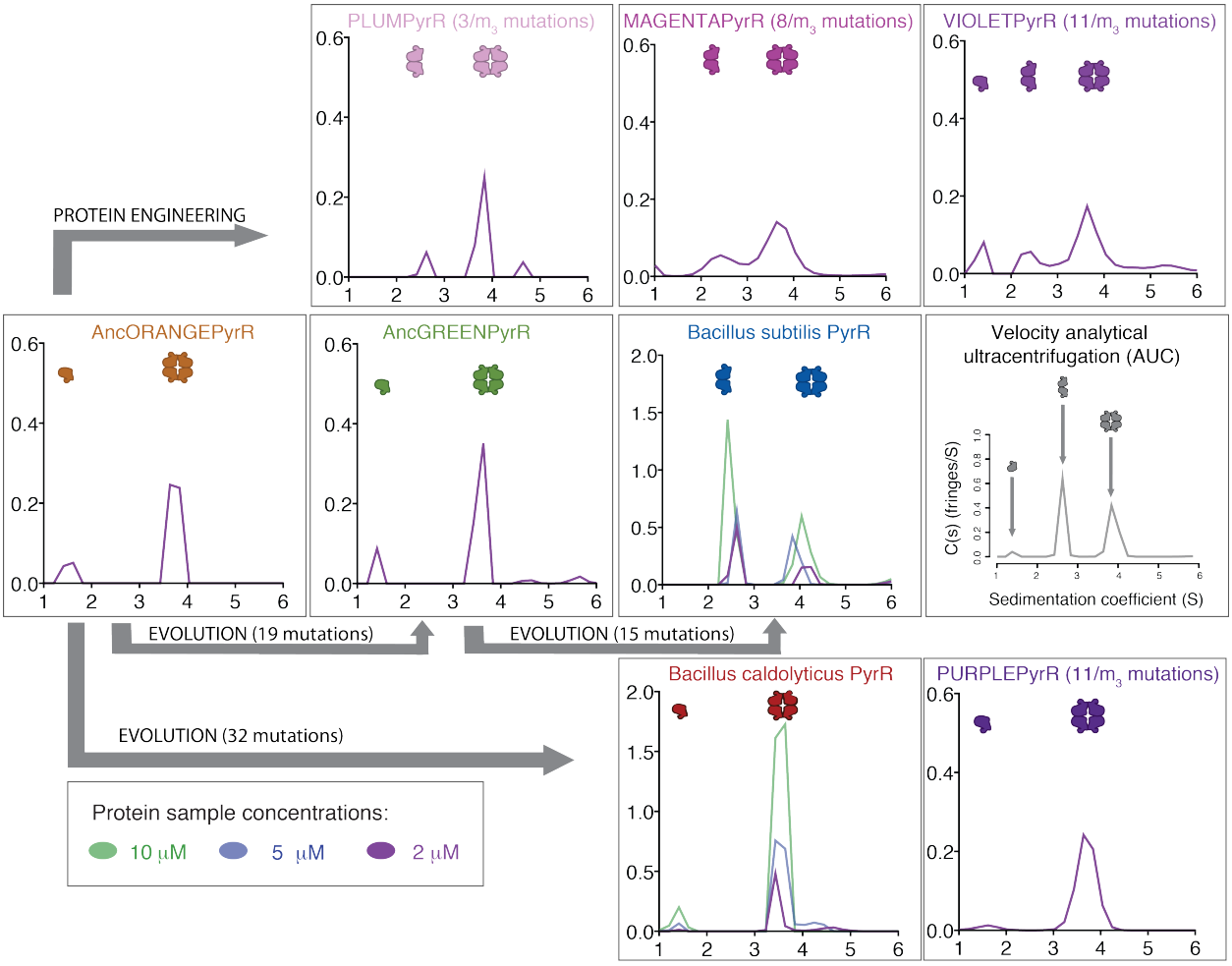


Figure S4 Analytical ultracentrifugation (AUC) analysis of the distribution of oligomeric species of PyrR proteins. We are showing only the c(s) distributions at 2 μM for the majority of samples for clarity of lowly populated species.

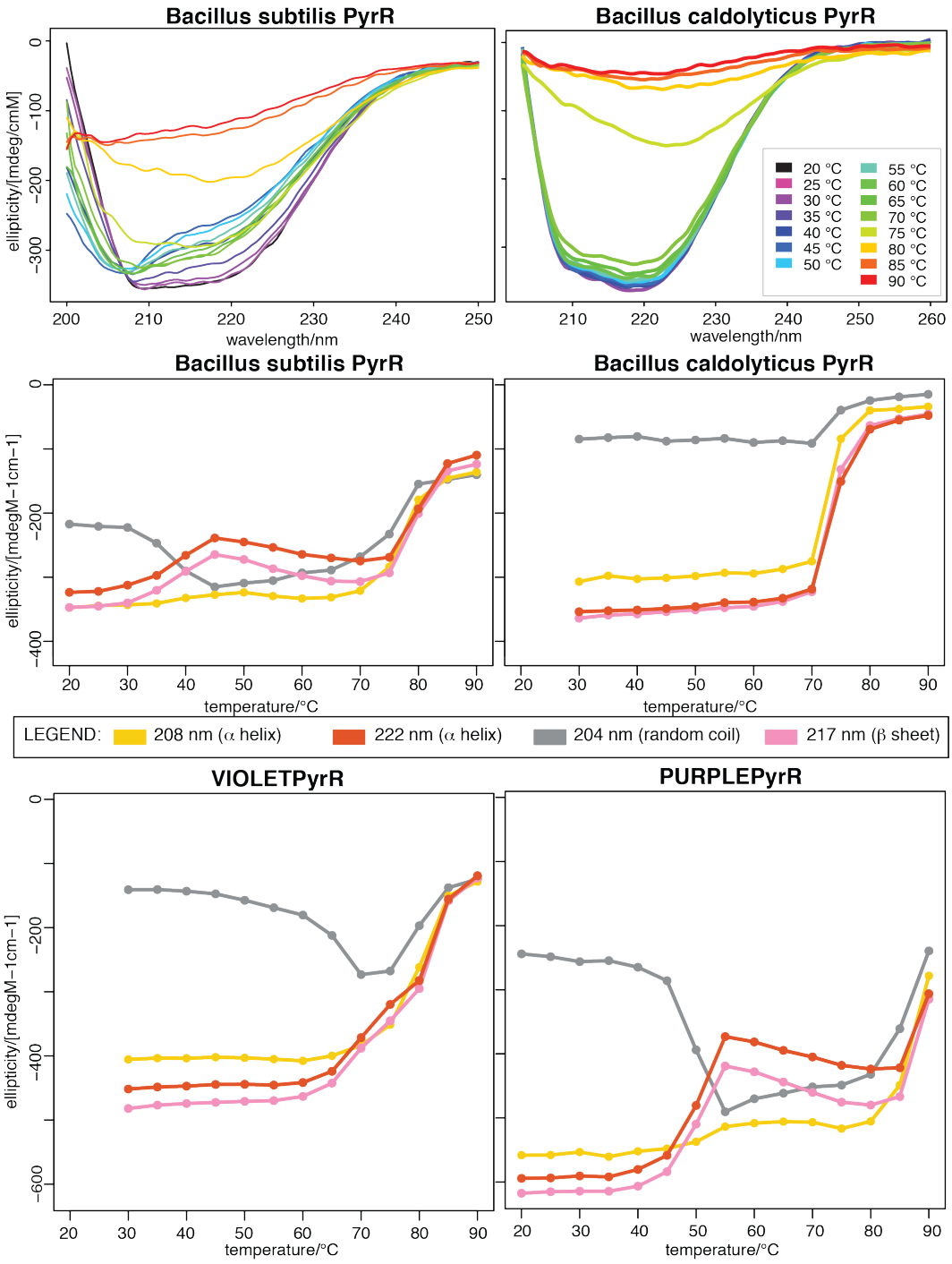


Figure S5 Circular dichroism (CD) spectra for PyrR proteins. CD spectra for *B. subtilis* PyrR and *B. caldolyticus* PyrR at a range of temperatures. *B. caldolyticus* PyrR unfolds cooperatively at approx. 75 °C, while the CD spectrum of *B. subtilis* PyrR changes differently for different wavelengths. Plotting change in ellipticity with temperature for different wavelengths shows that

while BcPyrR and VIOLETPyrR lose ellipticity for all wavelengths cooperatively, BsPyrR and PURPLEPyrR only partially lose ellipticity albeit at lower temperatures.

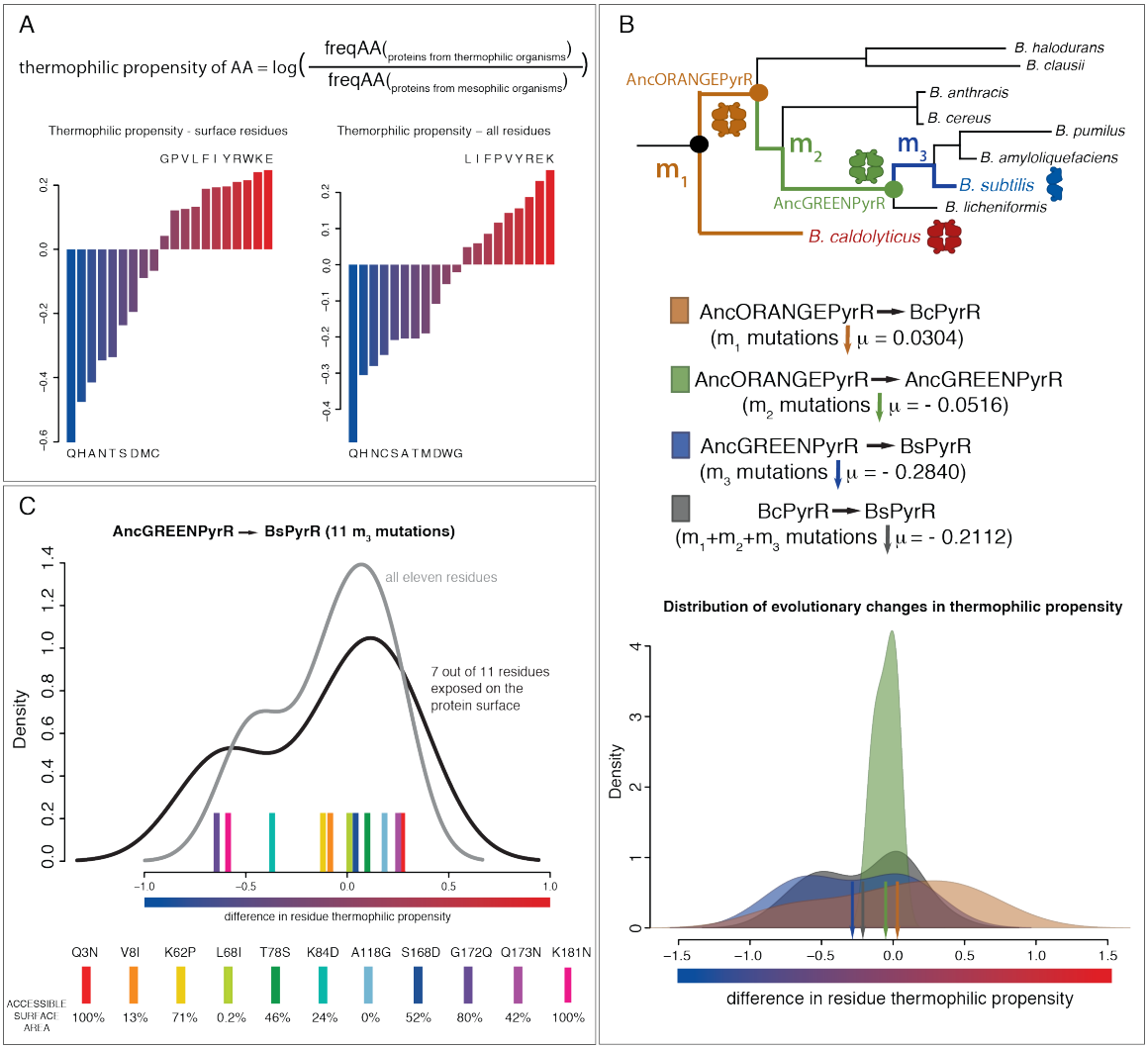


Figure S6 Evolutionary change in thermophilic propensity in the PyrR family. (A) Thermophilic propensity of a residue is defined as a log ratio of amino acid frequencies in proteins from thermophilic versus proteins from mesophilic organisms. Amino acid frequencies in thermophilic and mesophilic organisms were calculated in (*21*). (B) Thermophilic propensity increases in the PyrR family towards the thermophilic BcPyrR and decreases towards the

mesophilic BsPyrR. (C) Eleven allosteric mutations contribute significantly to the decrease in thermophilic propensity between AncGREENPyrR and BsPyrR. According to the calculated thermophilic propensity, mutations in three surface residues (G172Q, K181N and K84D) contribute most towards the decrease in thermostability.

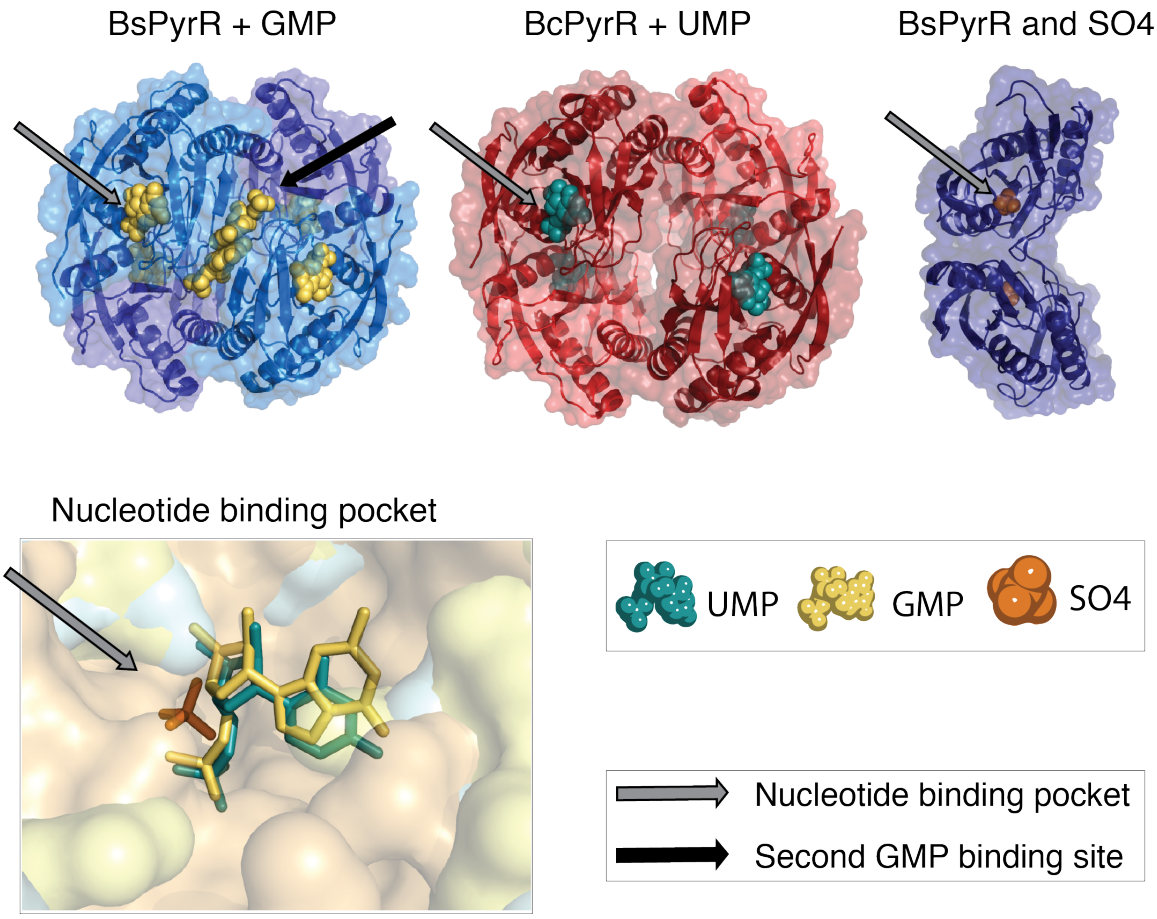


Figure S7 Binding of nucleotides to PyrR. UMP and GMP both bind to the same *nucleotide binding pocket*. Our crystal structures also show a second GMP binding site, stacking between the subunits. Most X-ray crystal structures of PyrR homologs have a sulphate ion (SO4) bound in the nucleotide pocket, if crystallised without the nucleotides. The binding site of UMP and GMP overlap very well, and SO4 binds close to the phosphate of the nucleotides.

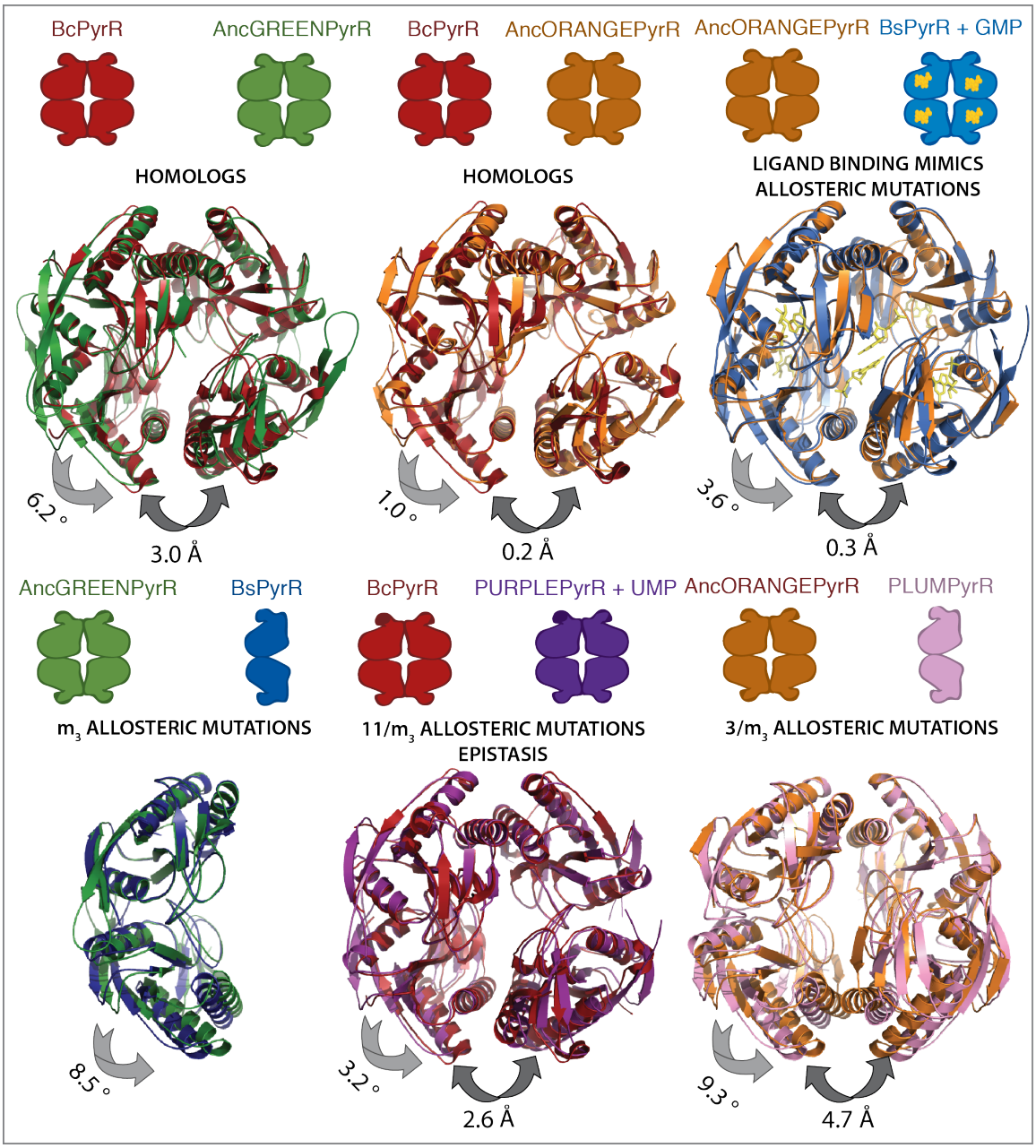


Figure S8 Inter-subunit geometry structural comparisons of extant, ancestral and engineered PyrR proteins. Although there are 23 substitutions between AncORANGEPyrR and BcPyrR (which is almost half of the mutations between BsPyrR and BcPyrR), the two proteins, have very similar inter-subunit geometries. The further 19 substitutions, from AncORANGEPyrR and AncGREENPyrR introduce more variation in the inter-subunit geometry, as well as a slight shift

in the elution peak in SEC MALS (Figure 2). Also shown is AncGREENPyrR, crystallised in a dimeric crystal form. Its inter-subunit geometry however, still exhibits an 8° rotation difference from BsPyrR, and structural superposition of AncGREENPyrR with the tetrameric BcPyrR shows that AncGREENPyrR inter-subunit geometry is compatible with tetramer formation, unlike that of BsPyrR. The engineered dimeric VIOLETPyrR has a very similar inter-subunit geometry to BsPyrR, incompatible with tetramer formation.

It is important to note that the packing in different crystal forms is not sufficient to explain the differences in the inter-subunit geometry we observe. All PyrR dimers crystalized in a C121 (or I121) form, but so did AncGREENPyrR and the tetrameric BsPyrR with GMP. Also, in our previous work we controlled for the variation due to the crystal packing in all the families analysed, including PyrR (*5*). For all the cases where a change in inter-subunit geometry could explain the difference in oligomeric state, the differences in inter-subunit geometry between dimers and tetramers was larger than the inter-subunit geometry between different crystal forms.

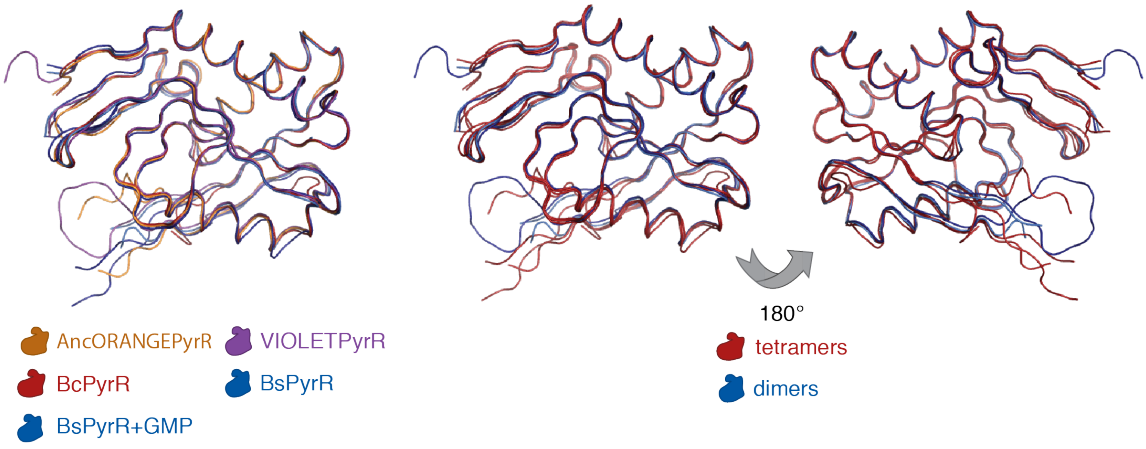
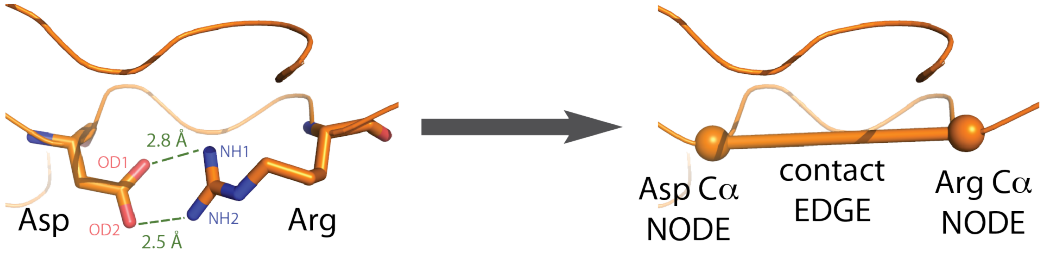
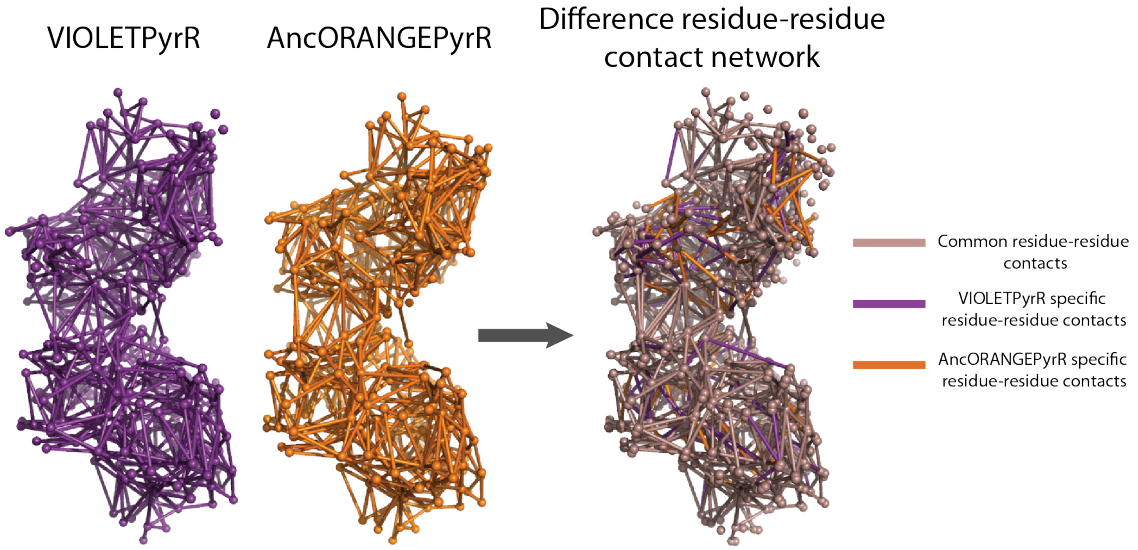


Figure S9 Structural backbone superpositions of single subunits of different extant, ancestral and engineered PyrR proteins. All versions of the individual subunits superpose very well, illustrating the close similarity between their individual folds.

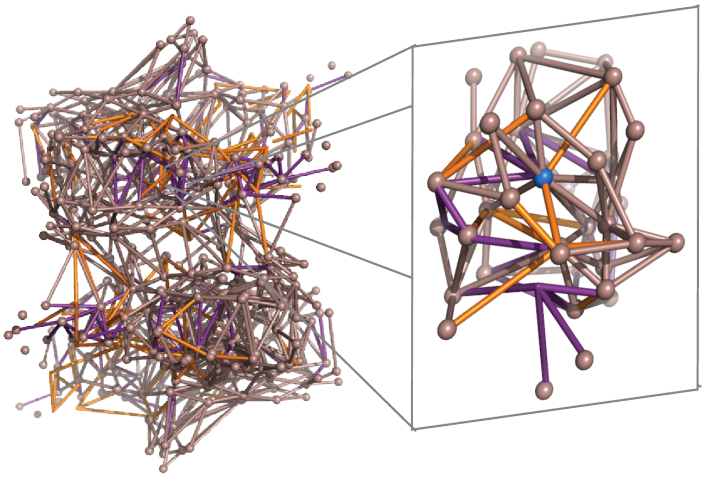
1.) Reduce the structure to a residue-residue contact network



2.) Compare the networks of two or more structures



3.) Define the extent of structural change around each residue of interest



For each residue make a subnetwork of radius = 2, which includes all the residues the residue of interest (in blue) is in contact with (radius = 1), as well as all the residues they are in contact with (radius = 2).
In that subnetwork, centred around the residue of interest, we count the number of structure specific contacts (violet and orange edges).
This number of different (violet and orange) contacts is not normalised for the overall number of contacts in order to preserve the information on the connectivity of the residue of interest.

Figure S10 Method for comparing structures using residue-residue networks.

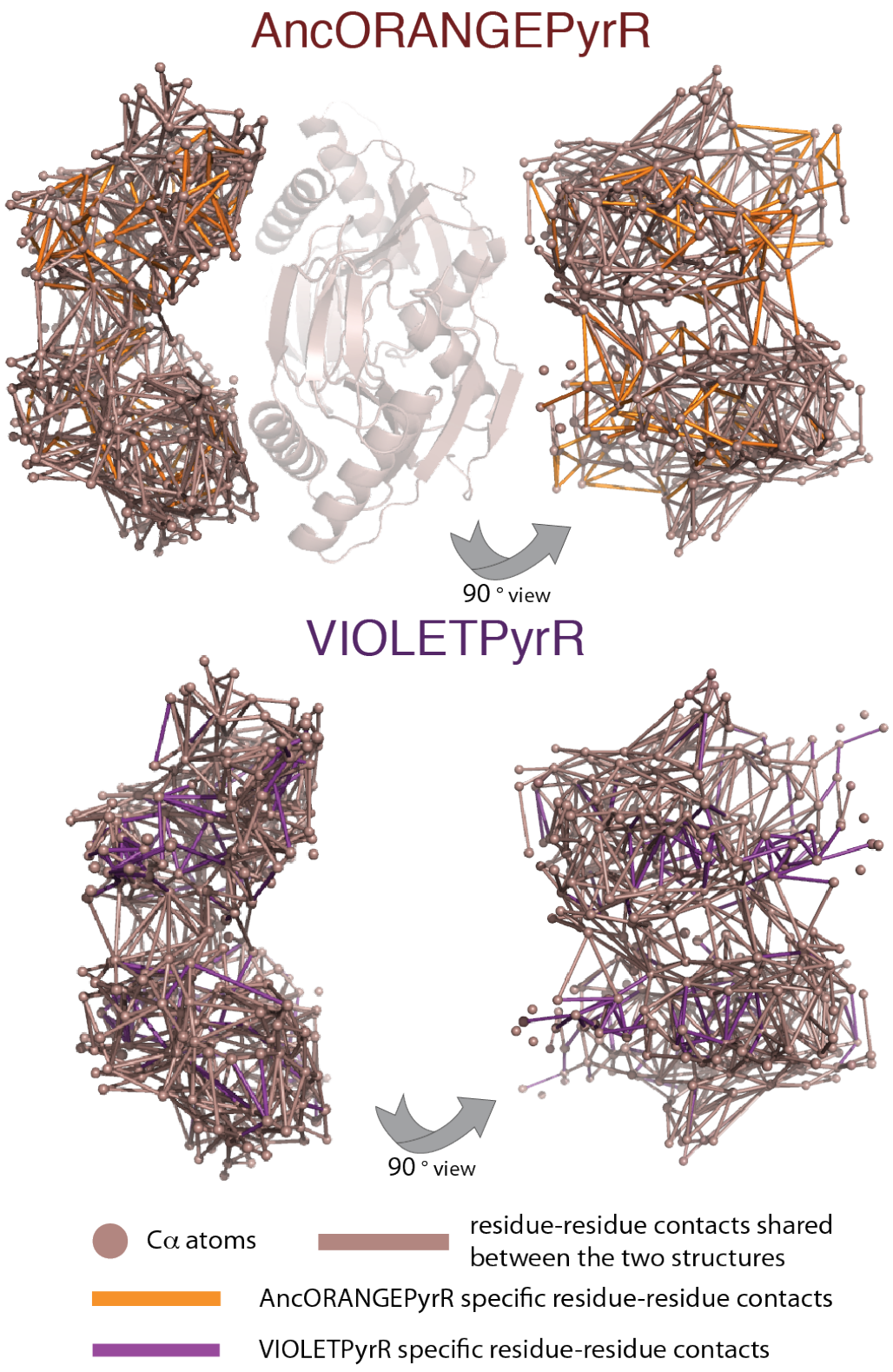


Figure S11 Residue-residue contact network differences between AncORANGEPyrR and VIOLETPyrR.

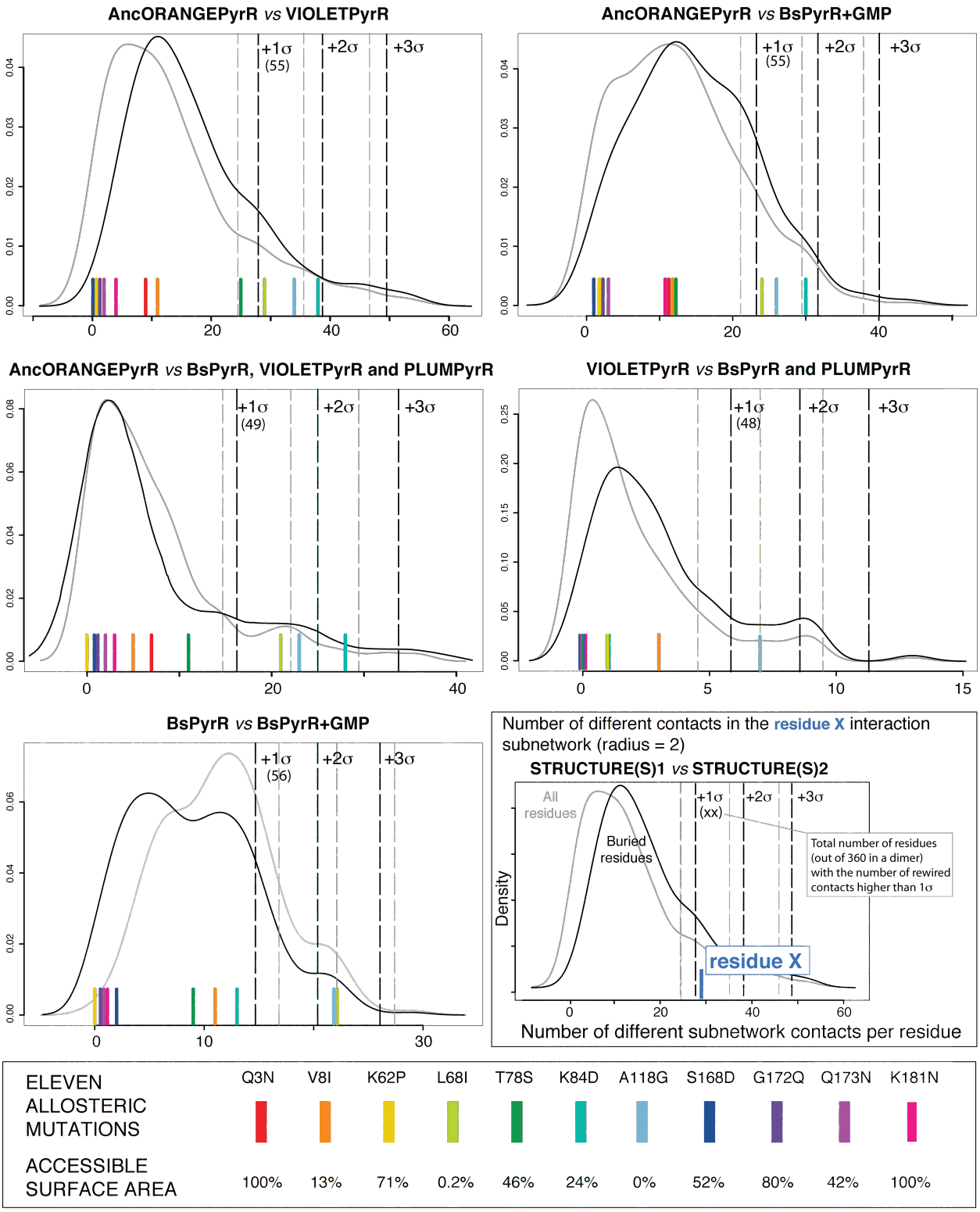


Figure S12 Eleven allosteric mutations and structural changes. For each residue in the dimeric PyrR structures we calculated the number of rewired contact in its subnetwork (radius 2, see

Figure S10). The kernel density plot shows the distribution of the number of rewired contacts for all ~360 residues of the dimer (gray distribution), and the distribution for only residues buried in the protein interior (plotted in black). The distributions change depending on which sets of structures are compared, for example, comparison of dimers (VIOLETPyrR with PLUMPyrR and BsPyrR) shows on average much less contact residue-residue rewiring than the comparison of VIOLETPyrR and AncORANGEPyrR. 48 to 56 out of 360 (13-16 %) residues show more than one standard deviation contact rewirings than an average residue (number shown in parentheses under the 1σ label) and thus form the region that undergoes significant structural change. Out of the residues changed by the eleven allosteric mutations, residues K/D 84, L/I 68 and A/G 118 fall into this group.

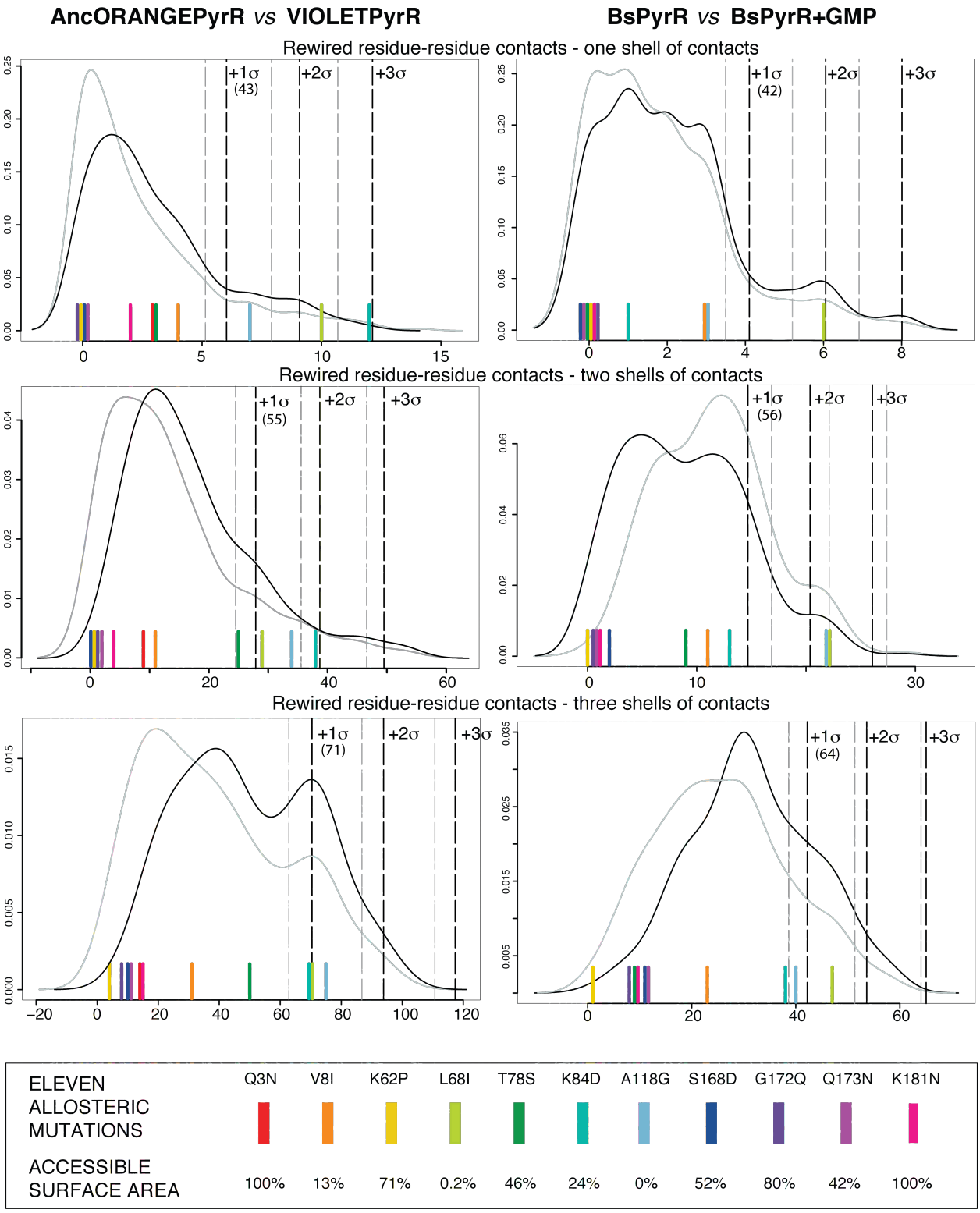


Figure S13 Comparison of residue-residue contact rewirings the eleven allosteric mutations are involved in for three levels of residue-residue contact shells. One shell of contacts considers only

residues that are in direct contact. Three out of the eleven allosteric residues (K/D 84, L/I 68, and A/G 118) show significant level of residue-residue contact rewiring when considering both one and two shells of contacts.

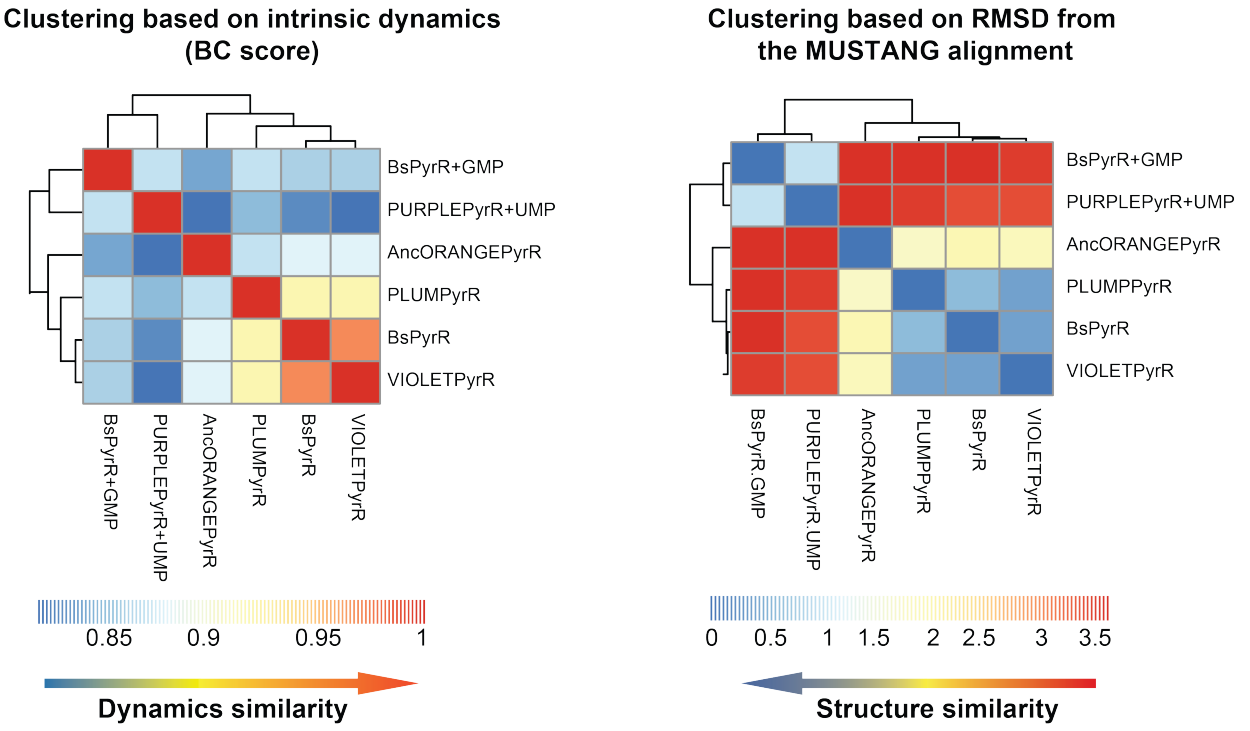


Figure S14 Hierarchical clustering of PyrR structures based on their intrinsic dynamics compared to the clustering based on RMSD. The left panel shows the clustering based on the intrinsic dynamics as quantified by the BC score (refer to Figure 5A in the main text), where high BC score denotes higher similarity in intrinsic dynamics. The right panel shows the RMSD obtained from the multiple structure alignment performed to determine the corresponding Cα atoms between the structures in Cartesian coordinate space, where low RMSD denotes high structural similarity. Both measures agree well with each other, even though they are comparing different properties. The BC provides a pairwise comparison of the covariances calculated from the normal mode vectors of each structure, from the regions that correspond in all of the structures,

as they encode the changes in dynamics that are conferred by the static change in atomic positions quantified by the RMSD.

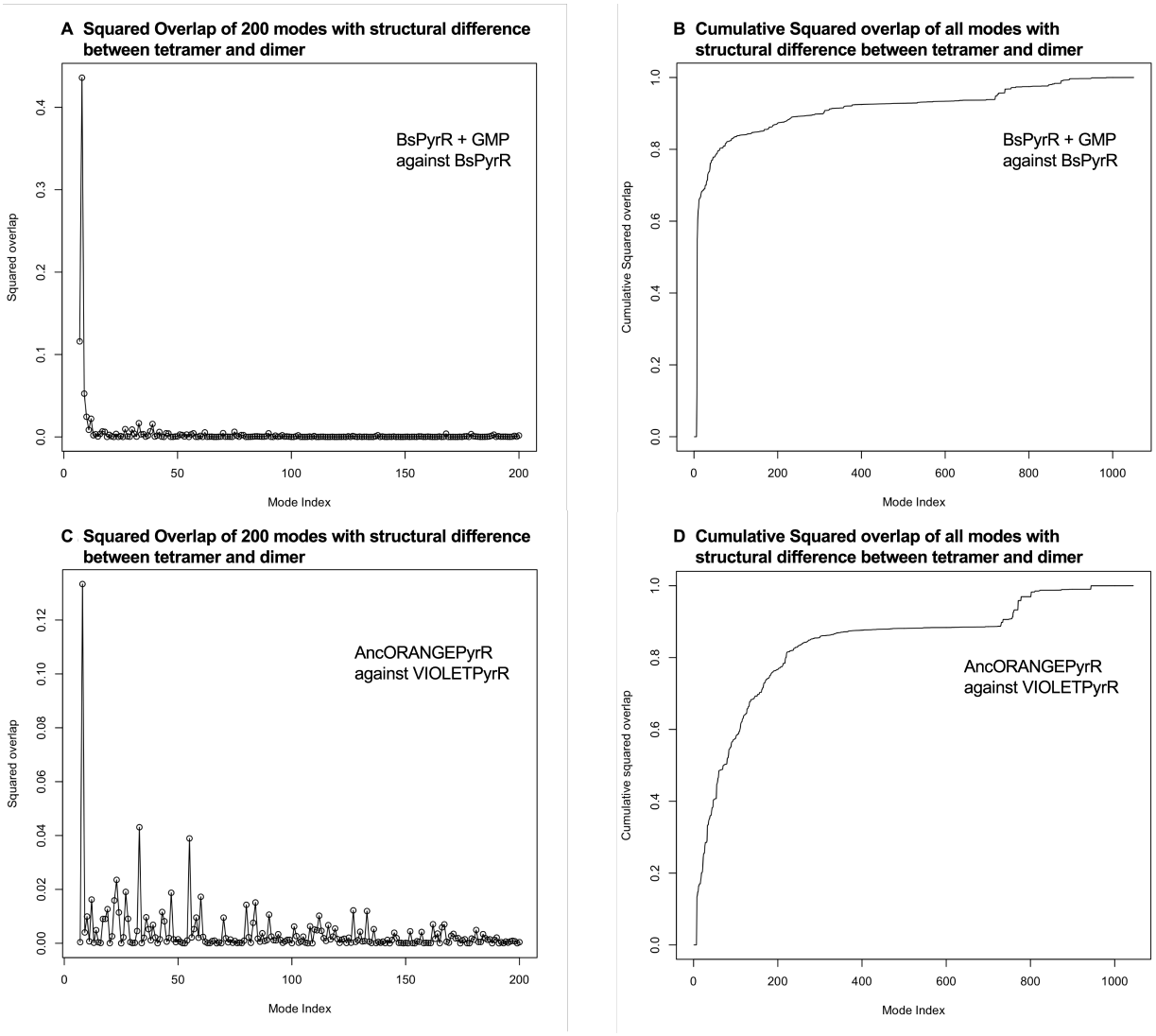


Figure S15 Overlap between the conformational transition from tetrameric to dimeric state, and the calculated normal modes. The left column (A and C) corresponds to the squared overlap of the first two hundred normal modes of the dimeric units of the tetramers with the structural difference vectors between the tetrameric and dimeric conformations, while the right panel (B and D) are their respective cumulative squared overlap plots for all the normal modes calculated. (A) The overlap of BsPyrR + GMP with BsPyrR, shows that the second lowest energy normal

mode is the top contributing normal mode (0.44 overlap) to the transition between the two oligomeric states. (B) The cumulative plot of the BsPyrR + GMP and BsPyrR shows that close to 70% of this transition overlaps with the three lowest energy modes. (C) A similar trend is observed between AncORANGEPyrR and VIOLETPyrR, with the second lowest energy normal mode being the top contributor (0.14). (D) The cumulative overlap between AncORANGEPyrR and VIOLETPyrR shows a much gentler ascent, with more than half of the transition lying within the 100 lowest energy normal modes.

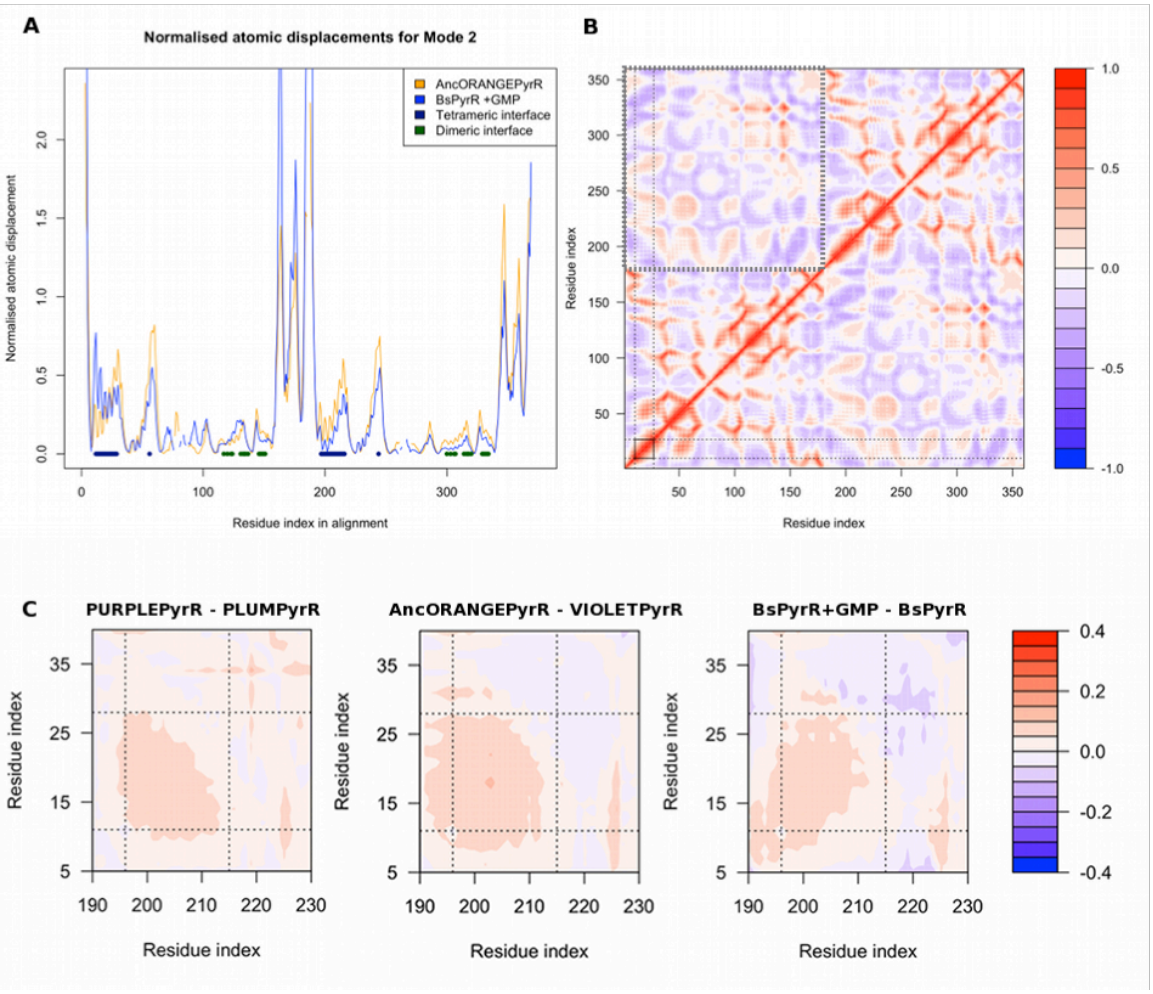


Figure S16 Dynamics of PyrR in evolution and function. (A) Normalised atomic displacement of Mode 2 of two tetrameric PyrRs: AncORANGEPyrR (orange) and BsPyrR + GMP (blue). The displacement is plotted as the residue index according to the structural alignment (x axis) versus the normalised atomic displacement (y axis). Videos S3 and S4 illustrate the corresponding structural change. The positions of the dimeric and the tetrameric interface in both subunits are indicated as dark blue and dark green dots, respectively. This low energy mode shows that the displacement of the tetrameric interface is greater than that of the dimeric interface, in both subunits. (B) Correlation matrix heatmap of BsPyrR. The motions of individual residues in the normal modes of proteins can range from being highly anti-correlated (-1, blue) to highly correlated (1, red), with 0 representing no correlation. Both the dark red and dark blue colours should be interpreted as regions with high correlations. We observe high intra- and inter-subunit correlations across the structure. The latter are highlighted by the large box on the top left part of the map. The tetrameric helix (small black box) correlates particularly well with other secondary structure elements, right up to the elements of the second subunit (highlighted with the dashed lines extended from the small square box). We infer that the changes in any of those parts of the structure could affect the tetrameric helices.

(C) Correlation difference heatmaps of tetrameric interface residues in dimeric halves of tetramers versus the dimers: PURPLEPyrR - PLUMPyrR+UMP, AncORANGEPyrR – VIOLETPyrR, and BsPyrR+GMP - BsPyrR. We observe higher correlations between the two tetrameric interfaces in tetramers than in the dimers. The pale red patch on the plot indicates the gain of above 0.1 in correlation in the tetramers, and includes the region corresponding to the tetrameric helix in both subunits. The tetrameric helix region is indicated by dotted lines at residue alignment positions, 12 – 29 (first monomer) and 197 – 216 (second monomer).

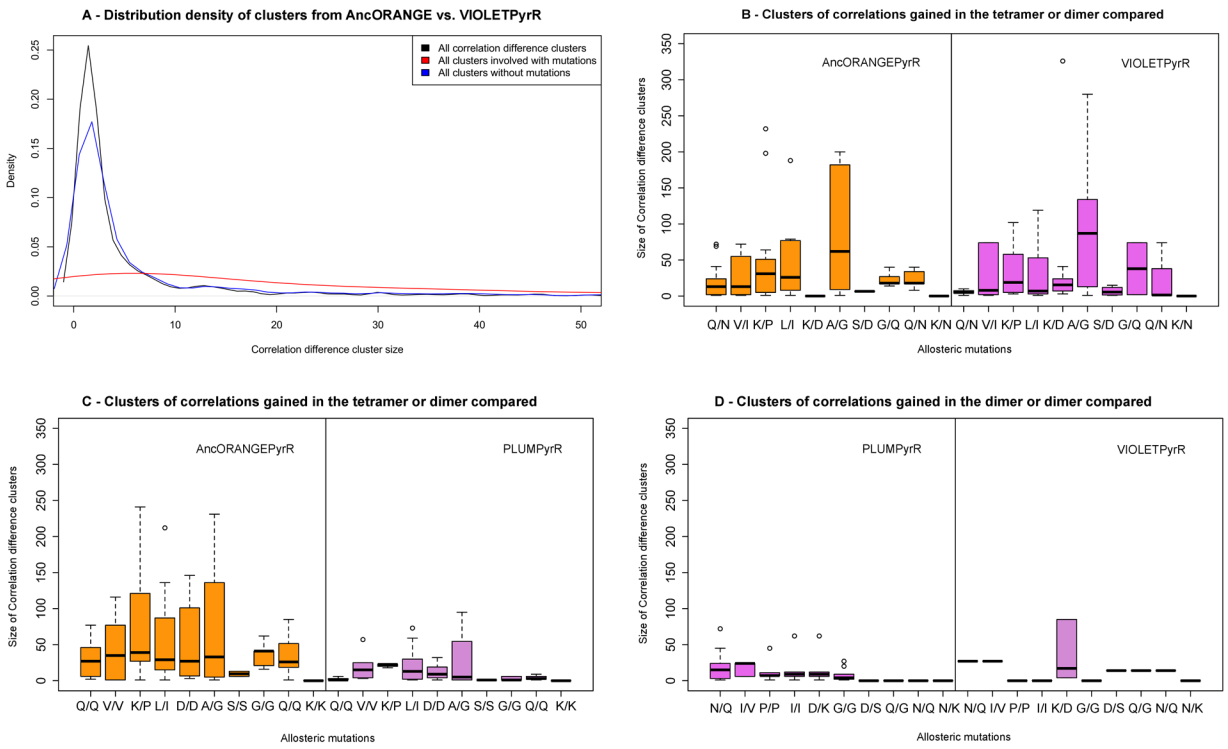


Figure S17 Statistical analysis of clusters of correlation differences between the dimeric units of AncORANGEPyrR and VIOLETPyrR (A). The clusters are chosen such that each possesses a minimum size of 1 pair of residues with a correlation difference score of or above +0.05, or below -0.05, reflecting a gain in correlation for the first and second structures, respectively. The kernel density plot shows the distributions of the correlation difference clusters in total (black), that involve the mutations (red) and without mutations (blue). The distribution shifts significantly when clusters of correlation differences that include the mutated amino acids are considered in relation to all the clusters collected in the analysis, where the peak of the mutations distribution increases to a cluster size of 7 amino acids. The box-and-whisker plots show the range of sizes (y-axis) of the clusters that reflect a gain in correlation in one structure compared to the other structure, for each of the mutated positions. (B) AncORANGEPyrR (left panel, orange) and VIOLETPyrR (right panel, violet), (C) AncORANGEPyrR (left panel, orange) and

PLUMPyrR (right panel, plum) and (D) VIOLETPyrR (left panel, violet) and PLUMPyrR (right panel, plum). The extreme outliers were excluded in all cases. Nevertheless, the positions Q/N and V/I in both (B) and the corresponding Q/Q and V/V positions in (C) are associated with the largest cluster (of size 628) due to their proximity to the tetrameric interface, as described by the large pink patches in Fig. S16C. The tetramer, AncORANGEPyrR gains a greater number of correlation difference clusters that are larger in size than the dimers, VIOLETPyrR and PLUMPyrR, with significant contributions from the three mutated positions, K/P, L/I and A/G. The difference between the dimers VIOLETPyrR and PLUMPyrR exhibits smaller clusters compared to the difference between AncORANGEPyrR and the dimers. PLUMPyrR is consistently implicated with smaller clusters of correlation gain when compared to AncORANGEPyrR.

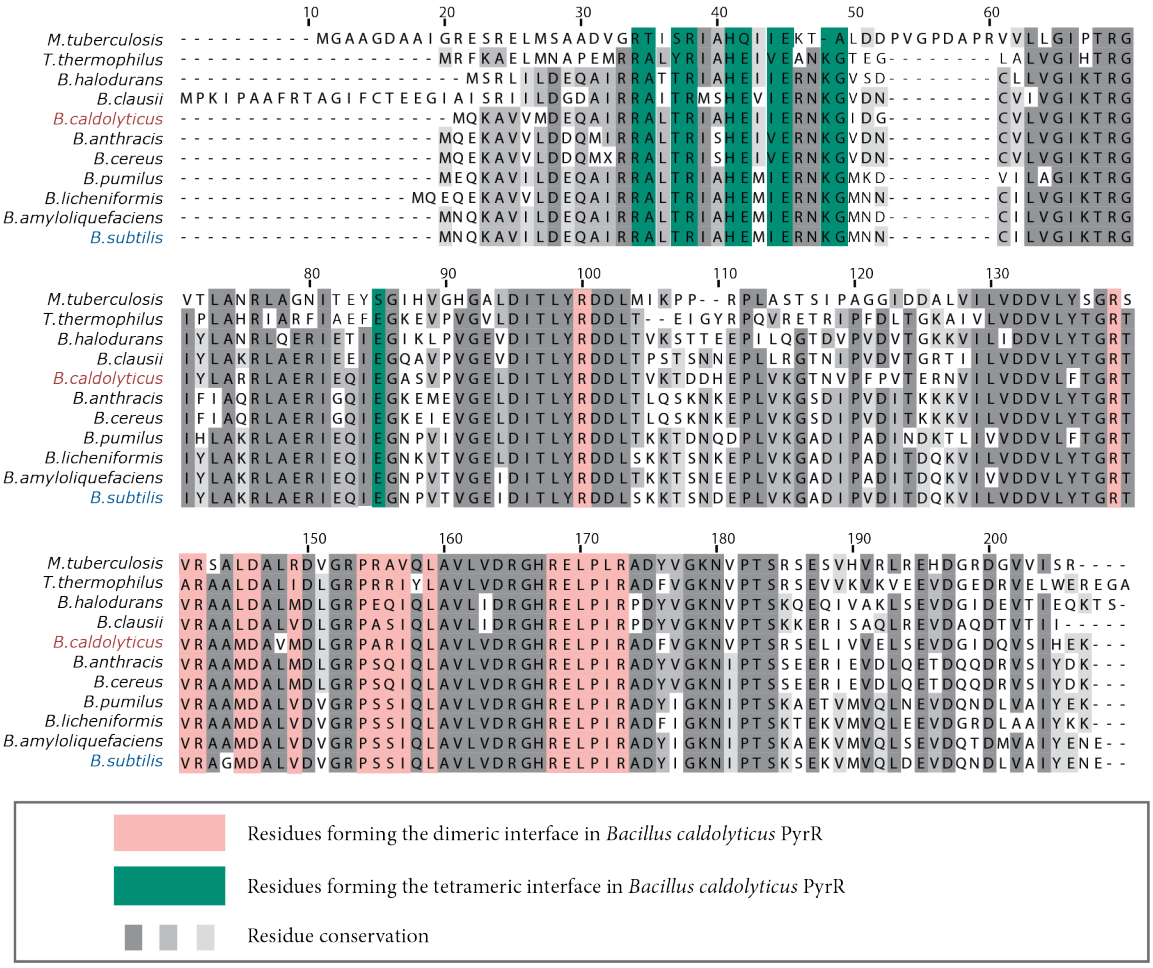


Figure S18 Multiple sequence alignment produced by MUSCLE (*37*). We reconstructed ancestral sequences based on this alignment with MrBayes (version 3.1) (*38*).

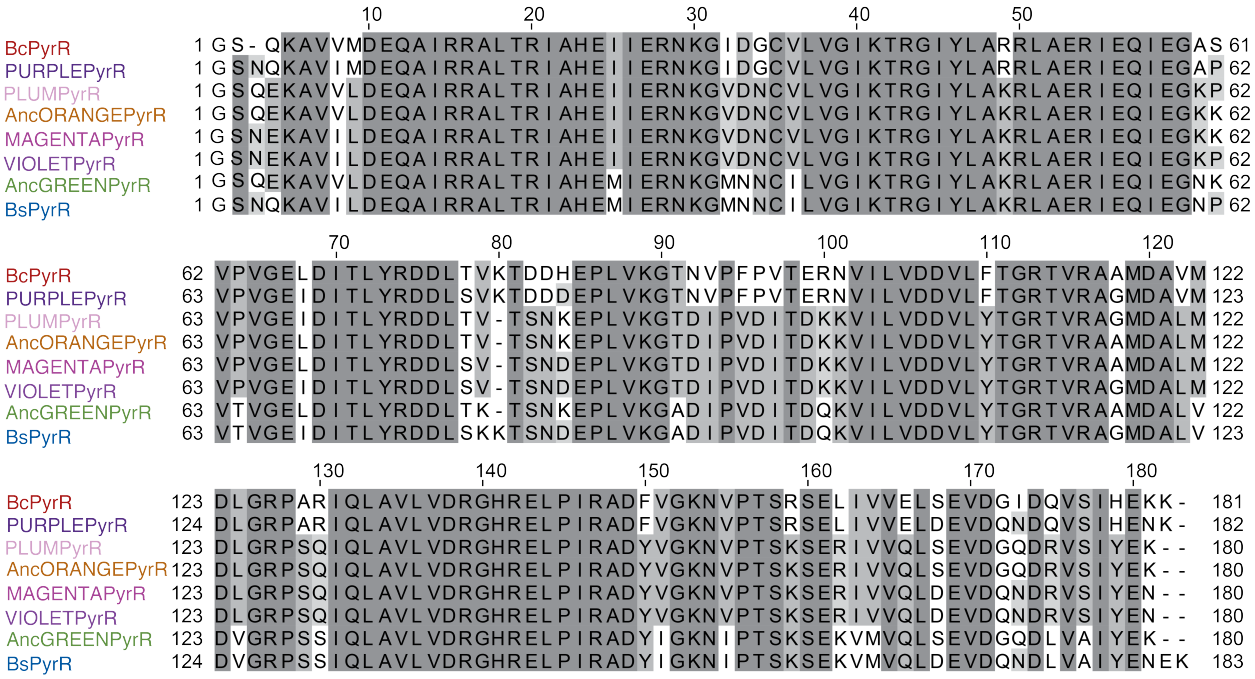


Figure S19 Multiple protein sequence alignment of BsPyrR, BcPyrR and all inferred ancestral as well as engineered PyrR proteins described in this study. PyrR proteins start with a Met residue, but due to the His-tag purification and His-tag cleavage with Thrombin, all the biophysical and structural experiments were performed with the PyrR proteins having a Gly-Ser instead of a Met in the beginning of the sequence. Residue numbers in all the figures and throughout the text as well as in the files in the PDB Database are according to this alignment.

Table S1 Sedimentation coefficients s(20, w) from velocity sedimentation AUC experiments. The s(20, w) value is a sedimentation coefficient corrected for viscosity and density of the solvent, relative to that of water at 20 °C. Values are shown as means with their standard deviations.

Protein	**Monomer**	**Dimer**	**Tetramer**
BcPyrR	2.0 ± 0.1		4.8 ± 0.1
BsPyrR		3.4 ± 0.1	5.3 ± 0.2
AncORANGEPyrR	2.2 ± 0.2		5.0 ± 0.3
AncGREENPyrR	2.2 ± 0.2		5.0 ± 0.1
VIOLETPyrR	2.0 ± 0.2	3.4 ± 0.1	5.3 ± 0.1
PURPLEPyrR	2.0 ± 0.1		3.7 ± 0.1
PLUMPyrR		3.1 ± 0.1	5.2 ± 0.1
MAGENTAPyrR		3.4 ± 0.1	5.0 ± 0.1

Table S2 Crystallographic data collection and refinement statistics

Protein	AncORANGEPyrR	AncGREENPyrR	PLUMPyrR	PURPLEPyrR+UMP
Space group	P 21 21 21	C 1 2 1	C 1 2 1	P 43 21 2
Cell Dimensions				
a, b, c (Å)	59.9, 102.3, 107.7	123.3, 67.0, 56.8	76.0, 57.2, 54.0	77.2, 77.2, 286.8
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 101.1, 90.0	90.0, 127.5, 90.0	90.0, 90.0, 90.0
Molecules/asymmetric unit	4	2	1	4
Data Collection				
Wavelength (Å)	0.9795	0.9200	0.9200	0.9200
Resolution (Å)	38.90 - 1.80 (1.84 - 1.80)	20.27 - 1.60 (1.63 - 1.60)	42.86 - 1.70 (1.74 - 1.70)	53.59 - 2.50 (2.61 - 2.50)
R_{merge}[a]	0.077 (0.646)	0.127 (0.479)	0.093 (0.645)	0.167 (1.166)
I/σI	7.6 (1.6)	5.6 (2.4)	11.5 (2.5)	10.9 (2.4)
Completeness (%)	94.8 (96.0)	95.4 (97.2)	99.0 (99.6)	99.0 (99.8)
Multiplicity	3.7 (3.8)	2.9 (3.0)	6.7 (6.9)	11.9 (12.1)
Refinement				
Number of unique reflections	55,721	54,426	19,065	29,162
R_{work}/R_{free}(%)	18.4 / 22.4	21.6 / 24.8	16.7 / 21.2	20.2 / 25.1
Average B factors (Å^2)	28.1	9.8	28.5	50.0
Number of atoms (nonhydrogen)	5,724	3,003	1,561	5,479
Stereochemistry				
Rmsd bond lengths (Å)[b]	0.018	0.023	0.019	0.015
Rmsd bond angles (°)[b]	1.82	2.25	1.94	1.87
Ramachandran outliers (%)[c]	0	0	0	0
Ramachandran favored (%)[c]	98.3	96.7	95.5	97.1
PDB ID	4P81	4P80	4P3K	4P83

Table S2 continued

Protein	VIOLETPyrR	BsPyrR+GMP	BsPyrR+SO4
Space group	I 1 2 1	C 1 2 1	C 1 2 1
Cell Dimensions			
a, b, c (Å)	53.8, 56.4, 60.2	97.9, 77.5, 99.8	76.5, 57.6, 54.8
α, β, γ (°)	90.0, 97.3, 90.0	90.0, 102.1, 90.0	90.0, 128.4, 90.0
Molecules/asymmetric unit	1	4	1
Data Collection			
Wavelength (Å)	0.9795	0.9795	0.9200
Resolution (Å)	42.55 - 2.20 (2.27 - 2.20)	54.58 - 1.93 (1.98 - 1.93)	42.93 - 1.30 (1.33 - 1.30)
R_{merge}[a]	0.208 (0.382)	0.043 (00.334)	0.059 (0.453)
I/σI	3.1 (2.2)	10.9 (2.4)	9.5 (2.3)
Completeness (%)	98.4 (98.0)	96.4 (96.9)	98.2 (96.1)
Multiplicity	2.6 (2.7)	2.7 (2.9)	3.4 (3.4)
Refinement			
Number of unique reflections	8,535	50,228	42,730
R_{work}/R_{free}(%)	20.8 / 27.0	18.1 / 21.4	18.0 / 20.6
Average B factors (Å^2)	25.2	45.1	19.3
Number of atoms (nonhydrogen)	1,603	5,845	1,649
Stereochemistry			
Rmsd bond lengths (Å)[b]	0.013	0.019	0.027
Rmsd bond angles (°)[b]	1.77	2.00	2.40
Ramachandran outliers (%)[c]	0	0	0
Ramachandran favored (%)[c]	97.8	97.2	97.9
PDB ID	4P84	4P86	4P82

a) Merging *R* factor

$$R_{merge} = \sum_{hkl}\sum_{i}\left|I_i(hkl) - \overline{I(hkl)}\right| / \sum_{hkl}\sum_{i} I_i(hkl)$$

b) Calculated in Refmac (*45*)

c) Calculated in Molprobity (*53*)

Supplementary Videos 1 and 2 – residue-residue contact networks

Supplementary Videos 1 and 2 show a 720° view of subunits A and B of AncORANGEPyrR and VIOLETPyrR represented as residue-residue contact networks. Contacts conserved between these two crystal structures are shown in pale violet, and the ones specific for AncORANGEpyrR and VIOLETPyrR in orange and violet, respectively. For orientation, the subunits C and D of the AncORANGEPyrR tetramer are shown in a cartoon representation.

Supplementary Videos 3, 4 and 5 – normal mode analysis

Applying Mode 2 to the trace of the X-ray structures of AncORANGEPyrR and of the BsPyR + GMP complex, we generated 50 conformations along Mode 2 to generate a movie illustrating the associated displacement. Note that the amplitude of the movement is arbitrarily chosen as the elastic network model predicts only the directions and not amplitudes of movements.

Video S3 – Animation of the AncORANGEPyrR dimeric units transformed along Mode 2. Mode 2 is the second lowest frequency mode obtained from AncORANGEPyrR, which was found to be the largest contributor to the conformational transition from tetramer to dimer in Fig. S14. The tetrameric helices (in cyan) are part of larger amplitude movements with respect to the rest of the structure, as also shown in the normalised atomic displacement plot in Figure S16A.

Video S4 – In this video the dynamics animation is obtained in the same way as in Video S3, but with both dimeric units of AncORANGEPyrR shown moving towards and away from each other as they transform along the mode, using 100 conformations for each unit.

Video S5 – Animation of the BsPyrR + GMP dimeric units transformed along Mode 2. Mode 2 is the second lowest frequency mode obtained from BsPyrR + GMP, which was found to be the largest contributor to the conformational transition from tetramer to dimer in Figure S15. The

tetrameric helices (in cyan) are among the regions displaced the most along this mode, as also shown in the normalized atomic displacement plot in Figure S16A.

Supplementary Methods

Ancestral sequence reconstruction

We performed the Bayesian inference by using MrBayes version 3.1 (*38*). The evolutionary tree topology, branch lengths and the sequences of ancestral nodes were calculated from a PyrR protein alignment using an estimated fixed-rate evolutionary model and PyrR from Mycobacterium tuberculosis for rooting the tree. MrBayes version 3.1. has nine default fixed rate models: Dayhoff (*54*), mtREV (*55*), MtMam (*56*), WAG (*57*), RtREV (*58*), CpREV (*59*), VT (*60*) and Blosum62 (*61*).In our analysis we have not determined the fixed rate model *a priori*, but have allowed MrBayes to jump between the models during the calculation until it converges with each of the models contributing in proportion to its posterior probability. For all the analyses, the ones estimating the evolutionary tree, as well as the ones calculating the ancestral sequences, the model with the highest posterior probability was Blosum62, followed by the WAG model. Each analysis was performed in four independent runs, with identical settings, and each run went through 1 000 000 generations, with the Markov Model Monte Carlo chain (MCMC) being sampled every 100 generations.

When inferring sequences of ancestral nodes, each node is calculated separately, as suggested by the authors of MrBayes (v. 3.1) (*38*), in order to integrate the uncertainty in the rest of the tree. The program provides the probability of each amino acid for each state (i.e. position in the sequence). The most correct way of inferring the ancestral sequence using the Bayesian principle would be to sample a variety of likely ancestral sequences based on the posterior probabilities for each position. However, as a series of low-throughput experiments had to be performed on each of the ancestral proteins, we chose the most likely sequence, i.e. a sequence where each position has an amino acid with the highest posterior probability, to test experimentally. All of the

ancestral sequences expressed in that way expressed and purified equally well as the extant proteins.

MrBayes (v 3.1) treats alignment gaps as missing data. In practice, this means the inferred sequence will be as long as the multiple sequence alignment used to infer it. MrBayes (v 3.1), therefore, implements a simple F81-like (*39*) model for binary type of data, which can be used to infer gap positions. The F81 model assumes all the sites in the sequence are independent, but the probability of change from *i* to *j* is proportional to the frequency of state *j*. In this simplified binary model, there are only two states, 1 representing a gap, and 0 representing an absence of a gap. The entire multiple sequence alignment was translated into a binary form (gaps (1), and absence of gaps (0)) and MrBayes (v3.1) was run for each node under the binary model. As in the case of inferred ancestral protein sequences, the program outputs a probability of a gap for each sequence position. We combined the binary and amino acid ancestral sequences in order to obtain the ancestral protein sequences of correct length.

Protein cloning, expression and purification

The cDNA of the wild type *B. subtilis* and *B. caldolyticus* PyrR was a gift from Prof. Robert L. Switzer from University of Illinois. Ancestral PyrR sequences were synthesized by GeneArt (Life Technologies), and we obtained the point mutants using the Quickchange protocol. We cloned all the constructs into a pRSET vector with a C-terminal His-tag and expressed in OverExpress™ C41 (DE3) cells. The lysis buffer contained 50 mM Tris pH 7.5, 300 mM NaCl, 1 M urea, and 2 mM β-mercaptoethanol. The proteins were eluted from the Ni-NTA column with 250 mM imidazole, and the His-tag was cleaved with Thrombin overnight at room temperature during dialysis into a buffer with 0.5 M urea (50 mM Tris pH 7.5, 300 mM NaCl and 5 mM β -mercaptoethanol). We then additionally purified the proteins on a size exclusion column (HiLoad

26/60 Superdex 200) and eluted them with a buffer containing 50 mM Tris pH 7.5, 150 mM NaCl and 1 mM DTT. Before setting up crystallization trays we additionally purified the proteins on a MonoQ column (GE Healthcare) and eluted them using a gradient of 50 mM to 1 M NaCl.

Analytical ultracentrifugation

Purified proteins were subjected to analytical ultracentrifugation using an Optima XL-I analytical ultracentrifuge (Beckmann) at various concentrations from 40 to 2 μM. Velocity sedimentation was carried out at 45,000 rpm at 10 ˚C in 50 mM Tris, pH 7.5, 150 mM NaCl, and 1 mM DTT using 12 mm double sector cells in an An60Ti rotor. The sedimentation coefficient distribution function, c(s), was analyzed using the Sedfit program, version 13.0 (*62*) with floated frictional ratios (f/f_o) of between 1.27-1.32. Masses of sedimenting species were calculated assuming a constant f/f_o. The partial-specific volume (v-bar), solvent density and viscosity were calculated using Sednterp (Dr. Thomas Laue, University of New Hampshire).

Crystalisation conditions

AncORANGEPyrR, PLUMPyrR, VIOLETPyrR and BsPyrR supplemented with 2 times excess GMPwere crystallized in Cryo 1 screen (Emerald BioStructures) condition 45, Cryo 2 screen (Emerald BioStructures) condition 5, Cryo2 screen condition 39, and CS Cryo (Hampton Research) condition 9, respectively. The crystals were flash-frozen by liquid nitrogen before data collection. BsPyrR was crystallised in the Wizard 2 screen (Emerald BioStructures) condition 38 and cryo-protected by soaking into the crystallization buffer supplemented with 25% glycerol before flash freezing. PURPLEPyrR supplemented with 1.2 excess UMP was crystallized in CS Lite screen (Hampton Research) condition 6 and cryo-protected by soaking into crystallization buffer supplemented by 30 % glycerol before flash freezing. AncGREENPyrR was crystallized in 1.2 M Ammonium Sulfate, 0.08 M Na Acetate pH4.8, 20% glycerol. Diffraction data were

collected at Diamond Light Source beam lines I02 and I04-1. AncORANGEPyrR, AncGREENPyrR, PURPLEPyrR, and VIOLETPyrR data were manually integrated with iMosflm (*63*) and scaled with Aimless (*64*). PLUMPyrR and BsPyrR data were processed by CCP4 (*42*), Pointless (*64*), Xds (*65*), and Xia2 (*66*). A structure of PyrR monomer (PDB ID: 1a3c) was used as a search model for molecular replacement by Phaser (*43*). The model was rebuilt using Coot (*44*) and the structure was refined by Refmac (*45*).

Thermophilic propensity

We used the frequencies of amino acids in proteins from mesophilic and thermophilic organisms from (*21*) and calculated the thermophilic propensity of an amino acid as its frequency in thermophilic proteins divided by its frequency in mesophilic proteins. In order to estimate the change of thermophilic propensity between two structures, for each mutation we subtracted the propensity of the amino acid in the second structure from that of the amino acid in the first. Fukuchi and Nishikawa (*21*) have shown that differences in propensity are more pronounced when only considering surface residues, rather than all the residues of the protein. We defined surface residues based on the percentage of accessible surface area of the residue in the structure, using thresholds defined by Levy (*67*).

Comparison of residue-residue contact networks

We define a residue-residue contact as a pair of residues in an X-ray crystal structure that have at least one pair of heavy atoms whose distance is less than the sum of their van der Waals radii (as defined in (*68*)), allowing for a 0.5 Å error to accommodate the uncertainties in atomic positions. Each residue is represented as a single node (illustrated by a Cα atom in Figures 4, 5, S10, and S11), but existence of a residue - residue contact was determined based on the distance of all heavy atoms, in both residue backbones and side chains. When we compared a pair of structures,

different contacts were those that existed between a pair of residues in one structure but not in the other. In cases where we compared sets of structures (in Figures S12 and S13) different contacts were the ones conserved in all structures from the first group and not existing in any of the structures from the second group.

To estimate the impact each individual mutation has on the residue-residue contact network (i.e. structure), we counted the number of different contacts from the first, first two, and first three shells of residue-residue contacts around the residue of interest (Figures S12 and S13). An average residue (when considering both common and different contacts between AncORANGEPyrR and VIOLETPyrR structures) has 8.7 ± 5.5 first shell contacts (residues making direct residue-residue contacts), 48 ± 26 first+second shell contacts, and 142 ± 65 first+second+third shell contacts. An average PyrR dimer between AncORANGEPyrR and VIOLETPyrR has 974 residue-residue contacts (or 443 per monomer). That means that three shells of contacts around an average residue cover a quarter of residue-residue contacts (142/443). At the same time, considering only the first shell of residue-residue contacts can be misleading, for example in the case of residues on the surface, or in weakly connected regions, that have more flexible side chains. They can rewire a relatively large number of contacts but this change might not propagate through the structure. We used the igraph package (http://igraph.sourceforge.net) for R for all network analyses.

Algorithm for identifying correlation differences between protein structures

We constructed correlation matrices from the normal modes as described by Ichiye and Karplus (*51*). The matrix elements have values between -1 and +1 for each pair of residues in a structure. Values of -1 and +1 indicate pairs with highly correlated displacements in opposite and parallel directions, respectively. A value of zero means no correlation. We compared the matrices of two

proteins by subtracting the absolute values of corresponding residue pairs according to the structural alignment. The resulting difference correlation matrices (ΔC) for two proteins inform us about gains and losses of correlations in one structure compared to the other. For example, when comparing a tetramer (T) with a dimer (D), D is subtracted from T (T-D); thus, positive values in the difference map correspond to gains of correlation in the tetramer and negative values to losses of correlation in the tetramer (all relative to the dimeric units).

The procedure used to define the clusters consists of parsing the ΔC matrix to find regions (a group of pairs of neigbouring residues) that undergo a gain or loss of correlations. When ΔC is plotted, these regions appear as red or blue 'patches'.

In practice, the search for clusters is performed iteratively as follows:

1) Starting from an amino acid of interest (e.g. a mutation point), one iterates over all its correlation values with other amino acids. On the plot of the ΔC matrix, it means starting from amino acid at position j on the Y-axis and moving along the X-axis through all the points that have (Y_j, $X_{(1\ to\ N)}$) as coordinates, N being the total number of amino acids.

2) A point (a pair of amino acids) with a value above the chosen threshold (for example, 0.1) is stored in a list, and the starting point moves by 1 position, along the X-axis to the right and to the left (if this point has not been visited before) and along the Y-axis (up and down). If the values of these four pairs are above the threshold, the pairs are stored in the cluster list and the search continues in all four directions from each of these.

3) The search grows subsequently, and stops in a given direction when a ΔC_{ij} score that is below the predefined threshold is reached. When the boundary of a given cluster has been explored, one continues iterating along the X-axis to find other clusters involving the

amino acid at position j (next red or blue ‘patch’ on the map). The coordinates to the clusters previously defined are stored to avoid visiting the same cluster more than once.

The clusters are selected for using two parameters, i) the minimum cluster size (number of points in a cluster), and ii) the minimum score. Only unique clusters, which differ in length (if the same points appear but includes more neighbouring points, then only the larger network is retained) and in the indexing of the points (for example, if a smaller cluster is sampled and these points differ from the ones collected before), are retained in this search.

The threshold value is chosen so that the statistical cluster analysis captures the correlation differences revealed by the calculation of the difference between the correlation maps of the two structures compared. For example, for the statistical analysis, the threshold at above +0.05, below -0.05 fully samples the pink patches that define the largest correlation difference in the tetrameric interface regions shown in Figure S16 C. In general, the differences in correlation usually do not exceed 0.15 in pairs of residues that are away from the diagonal of the correlation matrix, thus 0.05 is a reasonable cut-off for sampling them.

References

52. T. Pupko, R. E. Bell, I. Mayrose, F. Glaser, N. Ben-Tal, Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics (Oxford, England)* **18 Suppl 1**, S71 (2002).

53. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica Section D, Biological crystallography* **66**, 12 (Feb, 2010).

54. Schwartz, R.M., Dayhoff M.O., Chapter 22: A model of evolutionary change in proteins. *In Atlas of protein sequence and structure*, (1978).

55. J. Adachi, M. Hasegawa, Model of amino acid substitution in proteins encoded by mitochondrial DNA. *Journal of molecular evolution* **42**, 459 (May, 1996).

56. Y. Cao *et al.*, Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. *Journal of molecular evolution* **47**, 307 (Sep, 1998).

57. S. Whelan, N. Goldman, A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular biology and evolution* **18**, 691 (Jun, 2001).

58. M. W. Dimmic, J. S. Rest, D. P. Mindell, R. A. Goldstein, rtREV: an amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. *Journal of molecular evolution* **55**, 65 (Jul, 2002).

59. J. Adachi, P. J. Waddell, W. Martin, M. Hasegawa, Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *Journal of molecular evolution* **50**, 348 (May, 2000).

60. T. Müller, M. Vingron, Modeling amino acid replacement. *Journal of computational biology : a journal of computational molecular cell biology* **7**, 761 (2000).

61. S. Henikoff, J. G. Henikoff, Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 10915 (Nov 15, 1992).

62. P. Schuck, Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical journal* **78**, 1606 (Apr, 2000).

63. A. Leslie, H. R. Powell, Processing diffraction data with MOSFLM. *Evolving methods for macromolecular crystallography*, (2007).

64. P. Evans, Scaling and assessment of data quality. *Acta crystallographica Section D, Biological crystallography* **62**, 72 (Feb, 2006).

65. W. Kabsch, XDS. *Acta crystallographica Section D, Biological crystallography* **66**, 125 (Mar, 2010).

66. G. Winter, xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography* **43**, 186 (Dec 01, 2009).

67. E. D. Levy, A simple definition of structural regions in proteins and its use in analyzing interface evolution. *Journal Of Molecular Biology* **403**, 660 (Nov 05, 2010).

68. J. Tsai, R. Taylor, C. Chothia, M. Gerstein, The packing density in proteins: standard radii and volumes. *Journal Of Molecular Biology* **290**, 253 (Jul 02, 1999).