Computational Analysis of Dynamic Allostery and Control in the SARS-CoV-2 Main Protease

Igors Dubanevics[a,b,1] **and Tom C.B. McLeish**[b,2,*]

[a]School of Natural Sciences, University of York, UK; [b]Department of Physics, University of York, UK; [*]Corresponding author. E-mail: tom.mcleish@york.ac.uk

The dynamics and thermodynamics of the homodimer main protease protein SARS-CoV-2 are analysed at the level of Elastic Network Models (ENM). Correlated motions display a rich and heterogeneous dynamical structure, including allosterically correlated motions between the active sites in the two domains. Exhaustive 1-point and 2-point mutation maps, and their effect on binding free energies and allosteric free energies confirm the susceptibility identified in recent candidate inhibitor assays, but also suggest several new candidate regions for control and inhibition of the protease function, which may assist the development of current fragment binding screens.

1. Introduction

In 2019, a rapidly spreading disease named COVID-19 caused by the novel coronavirus SARS-CoV-2, has since generated a global pandemic. Preventive measures have been taken by a majority of countries, but no vaccine or anti-viral drugs exist, at the time of writing, although candidates are under trial. In the longer term, the identification of all potential inhibitor sites at all points of the viral life-cycle is of interest. Here we focus on the low-frequency dynamical structure of the virus' main protease, an important molecular machine in the viral cycle, and identify critical residues in its allosteric control. The work is informative for inhibitor design by identifying control regions of the protein that are distant from, rather than proximal to, its active sites. Allosteric mechanisms for distant control of binding and activation fall into two main classes: those which invoke significant conformational change (the original scenario of Monod, Wyman and Changeaux (1), and mechanisms that invoke the modification of thermal (entropic) fluctuations about a fixed, mean conformation (2–4). Such 'fluctuation allostery' recruits mostly global, low-frequency modes of internal protein motion, which are well-captured by correspondingly coarse-grained mechanical representations of the protein (5, 6). One effective tool at this level is the Elastic Network Model (ENM) (7). The ENM resolves protein structure at the level of alpha-carbon sites only, which are represented as nodes connected by harmonic springs within a fixed cut-off radius from each. Local point mutation can be modelled by changing the moduli of springs attached to the corresponding residue, and effector-binding by the addition of nodes and local harmonic potentials. The most significant contributions to the correlated dynamics of distant residues, and to the entropy of fluctuation come from global modes, whose ENM approximation allows straightforward calculation. This approach was successfully used to identify candidate control residues whose mutation may control allostery of effector binding in the homodimer transcription factor CAP (8). This study, and others, have shown that, while the huge reduction in the number of degrees of freedom that the ENM constitutes, does not capture the quantitative values of free energies, or their changes on mutation that are seen in experiment, it can identify the qualitative nature of the thermal dynamics of a protein. Furthermore,

it's coarse-graining can determine which residues present as candidates for allosteric control through mutation. The method, and the open software ('DDPT') used in the previous study on allosteric homodimers (9) is deployed here in a similar way (see Methods section) to a coarse-grained ENM model of the SARS-CoV-2 Main Protease.

A. The SARS-CoV-2 Main Protease Protein. At the time of this study (May 20, 2020) more than 16,000 papers have been published in relation to the virus (See COVID-19 Primer). However, work is still in progress to identify biological and molecular pathways the virus takes. Fortunately, significant research has already been directed to very similar coronavirus - SARS-CoV, first identified in 2003. The SARS-CoV and SARS-CoV-2 genetic sequences are almost 80% identical (10). Both viruses encode the main protease (M^{pro}), also known as the 3C-like protease ($3CL^{pro}$). In its active form M^{pro} is a two protomer homodimer with one active per the homodimer chain (11). Although M^{pro} is a relatively compact protein (less than 310 residues per chain), it plays a vital role in the viral cycle of both coronaviruses: it divides polyproteins expressed from the viral mRNA into its functional non-structural units (12). This functional role makes SARS-CoV-2 M^{pro} an appealing target for drug design. The major research effort to date has been focused on the competitive inhibition of SARS-CoV-2 M^{pro}, i.e. by directly targeting the active site with molecules that competitively bind to the active site "pockets" (10, 13–15). A significant body of work has been recently published investigating inhibitors for the SARS-CoV-2 main protease via virtual screening and MD simulations(16–23). In 2011 it was found that N214A mutation dynamically inactivates SARS-CoV M^{pro} (24). Due to the high sequence conservation between both coronaviruses, and 96% amino acid sequence identity between the main proteases, SARS-CoV-2 M^{pro} might posses similar allosteric features. Furthermore, in another study, researchers found the the root mean square deviation (RMSD) of 0.53 Å for apo (ligand-free) forms of two corona viruses' Main Proteases (PDB accesion 2BX4 and 6Y2E) (10). These evidence that the N214A mutation operates through a fluctuation allostery mechanism and structural similarities between two proteases motivates the analysis of the coarse-grained dynamic structure of SARS-CoV-2 M^{pro} reported here. We apply the ENM techniques of (8) to this purpose, looking in particular to identify non-active, yet allosteric, sites for non-competitive inhibition.

2. Simulations, Results and Discussions

No crystallographic structure of the SARS-CoV-2 M^{pro}* active form with a polyprotein is available to date. Only empty (apo) structures or structures with synthetic ligands/substrates attached to the active site available. Therefore, the ENM study reported here used a recent crystallographic structure (PDB accession 6LU7 (13)) with competitively inhibited active site to calculate fluctuation free energies and consequent allosteric energies and their modification under mutation. The inhibited (with N3 inhibitors, referred as "holo2") and ligand-free (apo) structures of the protein are almost identical shown in figure 1. Resolutions for the structures shown are 2.16 Å and 1.75 Å respectively, while RMSD between them is under 1.5 Å for C_α atoms. Evidently, very little structural change happens upon the inhibitor binding. These findings further support the hypothesis of dynamically driven allosteric control of SARS-CoV-2 M^{pro}, and provide a structure (6LU7) on which to base an ENM construction (**SI, Sec. A**).

The resulting ENM of M^{pro} is shown in figure 2. It takes $C_\alpha$node masses as the whole residue mass, and uses a cut-off distance for harmonic connecting springs of 8 Å, based on comparison of mode structures with full Molecular Dynamics simulations in previous work on Catabolite Activator Protein (8, 25)(**SI, Sec. B**). Balancing the requirements of: (i) sufficient spatial resolution of dynamics;

*From now on, M^{pro} will refer to the SARS-CoV-2 main protease protein

Fig. 1. Inhibited (green) and ligand-free (pink) SARS-CoV-2 M^{pro} crystallographic structures superimposed using the Combinatorial Extension (CE) algorithm in PyMOL (Schrödinger). The RMSD between two structures is 1.48 Å. Both proteins are shown as secondary structure cartoons while N3 inhibitor is shown as spheres. PDB accessions are 6LU7 and 6Y2E, respectively.

(ii) requirements not to include unphysically small-scale structure; (iii) acceptable convergence of thermodynamic calculations; (vi) compatibility with the previous studies (26) leads to the choice of summing the first real 25 modes in SARS-CoV-2M^{pro} ENM calculations (**SI, Sec. C**).

A. Residue-residue dynamic cross-correlation map. The first quantity of interest is the map of residue-residue cross-chain dynamic correlations, which indicates for each residue on the protein those other residues whose motion correlates with its own (**Eq. 1**). This gives both a detailed summary map of the homodimer dynamical structure, and is also significant thermodynamically since the same elastic communication drives both correlations and allosteric control (3). The dynamic cross-correlation of motion map for the all residues in the ENM apo (ligand-free), holo1 (only one active site at chain A occupied) and holo2 (both active sites occupied) structures are shown in figure 3. We discuss the dynamic features of each structure in the following:

A.1. apo. The N-terminus of each chain positively correlates with residues adjacent to the active site (100-200) on the other chain. This is due to physical proximity rather than allostery (**Fig. 3a**, wide black rectangles in lower right and upper left quadrants) . Significantly, the dynamics of active sites on both chains positively correlate, and allosterically, with each other (**Fig. 3a**, green square and rectangle in lower right quadrant). These regions are spatially far away: we can not see them at the corresponding location on the distance map (**Fig. 3a**, green square and rectangle in upper left quadrant). Also, an alpha helix (res 201-214) on one chain correlates positively with the active site on the opposite chain (**Fig. 3a**, narrow black rectangles in lower right and upper left quadrants). However, this correlation is partially caused by spatial proximity. Thus, we observe dynamic coupling on a distance that supports previous findings regarding SARS-CoV M^{pro} dynamically allosteric inactivation.

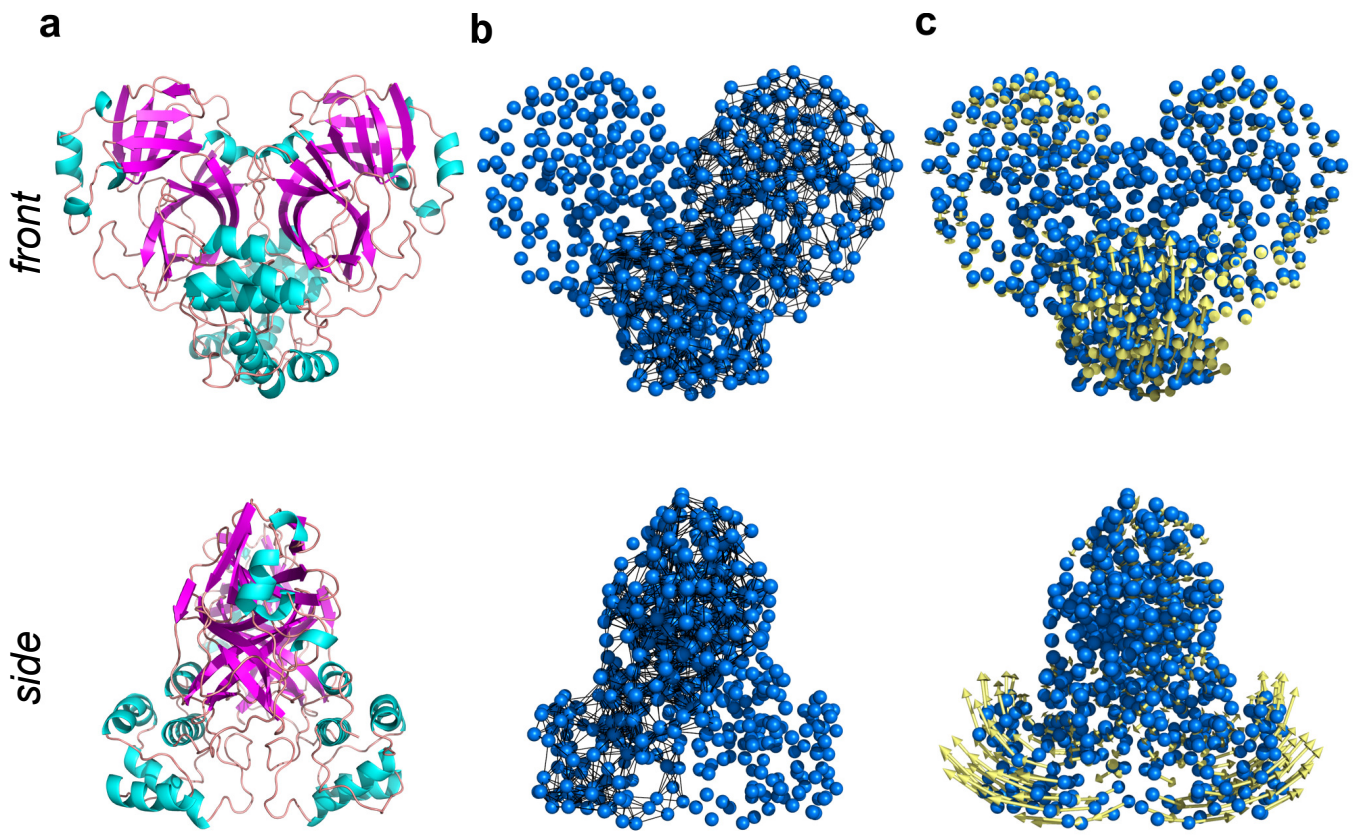

Fig. 2. Constructing ENM of SARS-CoV-2 M^{pro} step-by-step. **a**, SARS-CoV-2 M^{pro} secondary structure cartoon **b**, Elastic model of M^{pro} generated with PyANM package in PyMOL. C_α atoms are shown in blue; while node-connecting springs (black) are shown only for one chain for comparison. **c**, The first real vibrational mode eigenvectors (yellow) visualisation. For clarity, displacement vectors are scaled 5 times.

A.2. holo1. The positively correlated dynamics between the active sites are strongly enhanced by an addition of the first ligand, especially in regard to the loop between 150-160 (**Fig. 3b**, green square in the lower right quadrant). Interestingly, and by contrast, the region displaying positive correlation around res 50-70 in apo form is decreased in the holo1 structure (**Fig. 3b**, green rectangle in the lower right quadrant). Now, the structural symmetry is broken upon ligand addition to chain A, permitting the asymmetric magnitude of correlation between chains A and B (**Fig. 3**b, across the diagonal of the lower right quadrant). The biologically active residues (green ticks) show up in the ligand's correlation with (host) chain A. However, four other regions, not cited in the literature to date, also show strong dynamic correlation with the ligand. Two of them (res 17-32 and 120-131) can be vividly observed as spatially proximal to the ligand from the corresponding distance map. Two other regions exhibiting positive cross-correlation are, however, distant from the ligand (**Fig. 3b**, purple rectangles in lower right and upper left quadrants). These regions span res 67-75 and 77-91, respectively, and include two beta sheets in each case. The previously reported dynamically allosteric residue 214 (chain B) correlates positively with ligand on chain A; potentially due to spatial proximity. Moreover, the ligand shows positive correlation at distance with the beta sheets on the opposite chain (**Fig. 3b**, black rectangles in upper right quadrant). The fact that the ligand's motion positively correlates with beta sheets at res 67-75 and 77-91 on both chains suggests strong

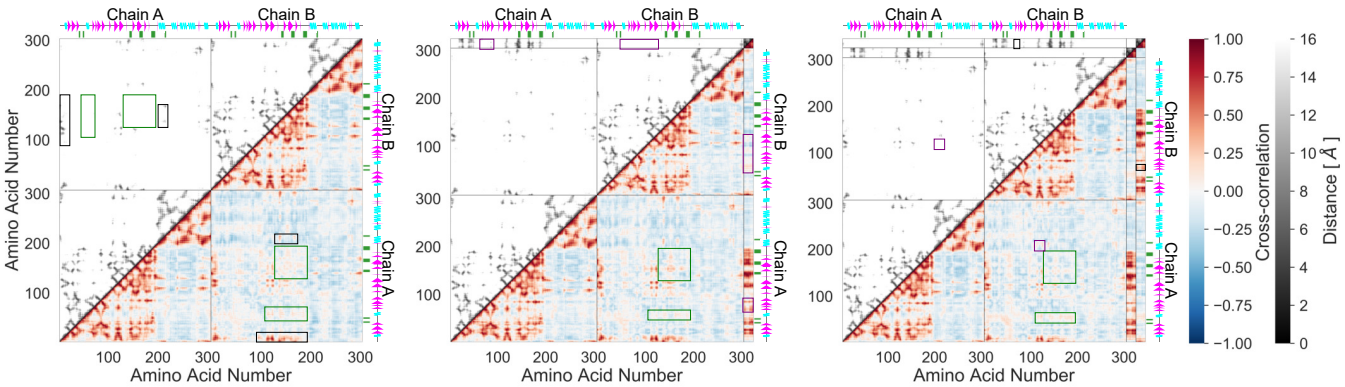

Fig. 3. The cross correlation of the motion for the first real 25 modes and distance between residues(C_α nodes) as shown in 2-dimensional residue space for (*left*) apo, (*middle*) holo1 and (*right*) holo2 forms of 6LU7 ENM . The first colour scale show the extent of cross correlation, with a cross correlation of 1 (red) indicating perfectly correlated motion, -1 (blue) showing perfectly anti-correlated motion and 0 (white) no correlation. The second colour scale (black to white) depicts the Euclidean distance between two C_α nodes in the Cartesian space in 0-16 Å range. The secondary structure of M^{pro} is indicated along the residue axes, with cyan waves indicating alpha helices, and magenta triangles indicating beta sheets. The green ticks on the axis indicate the location of the biologically active residues (**Tab. 1**). The cross correlation matrix was calculated using only the C_α atoms for the protein and all heavy atoms for ligand (N3 inhibitor).

chain-ligand coupling around res 67-91 region on both homodimer chains.

A.3. holo2. When a second ligand is added, the strong correlation between active sites present in holo1 is diminished (**Fig. 3c**, green square in lower right quadrant); while previously lowered correlation reaches the apo form level (green rectangle in the same quadrant and figure). Nevertheless, the cross chain coupling between active site and the alpha helix (201-214) is further increased. The correlation is split into three distinctive regions in locally wedge-like shapes on the map, two of which are around active site. These structures were also seen in the CAP cross-correlation map (8), and indicate the two contributing beta strands acting as a local hinge region. The third region is located around a beta sheet at res 120-130, which is not bound to the ligand (**Fig. 3c**, purple squares in lower right and upper left quadrants). The cross-correlation displays an interesting structure along the helix from residue 201 to 214: the third region is split into three zones: positive, negative and positive correlation. This sign change reminds 3-rd harmonic of a standing wave with two nodal points, whereas in 3D those are nodal planes. The negative correlation region is absorbed by the two positive regions as we reach residue 214. Unsurprisingly, the two ligands exhibit identical correlations with their host and opposite chains. Noteworthy, each ligand has dynamics correlating positively with the termini of the opposite chain, arising principally from the spatial proximity of C and N termini to the opposite chain's active site.

B. Mutation scans for thermodynamic control. The ENM calculations were extended to calculations of fluctuation (entropic) free energies (**Eq. 2**) of various modelled wild type and mutated state of SARS-CoV-2M^{pro}. 'Point mutations' are modelled in these ENM calculations, as in (8) by softening or stiffening all the harmonic springs attached to each residue in a complete scan of equally spaced spring moduli in range from 0.25 to 4.00. Figure 4 reports the free energy changes $(G_{mut} - G_{wt})/|G_{wt}|$ induced in the entire homodimer when this point mutation is done. Figure 5 reports the effect of the same mutational scan on the allosteric free energy K_2/K_1 (**Eq. 3**) of

Table 1. SARS-CoV-2 Main Protease key information used in this study: PDB ID; experimentally identified bioactive residues of SARS-CoV M^{pro} reported in literature; active regions which we identified in this study, distant form the active site for SARS-CoV-2 M^{pro}. Alpha helices and beta sheets are denoted in brackets as α and β, respectively.

Protein	PDB ID	Experimentally Identified Bioactive Residues	Computationally Identified Control Regions (this study)
SARS-CoV-2 Main Protease	6LU7	41, 49, 143-145, 163-167, 187-192 (17) 214 (24)	101-104 (β), 201-214 (α), 261-274 (α) 9, 12, 14, 276, 281, 284 (interface)

binding between the two active sites.

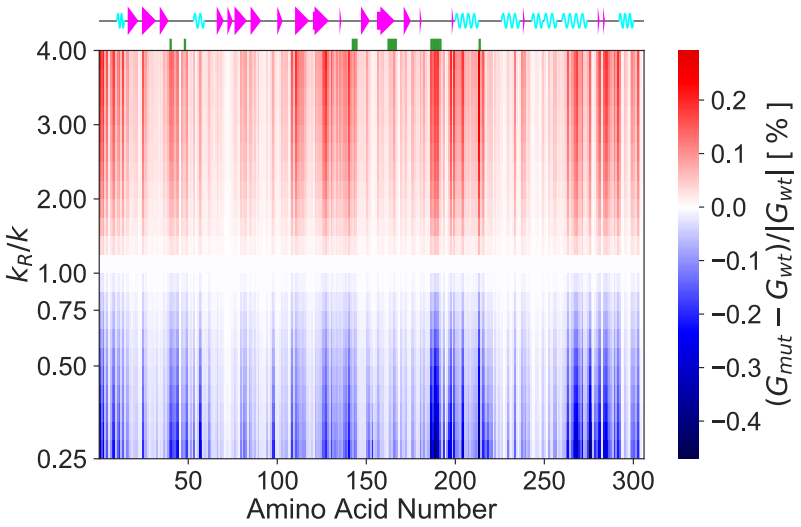

Fig. 4. A map for the vibrational free energy change in M^{pro} calculated from the ENM over 25 modes. The map plots the relative change in free energy to the wild type ($(G_{mut} - G_{wt})/|G_{wt}|$) due to the dimensionless change in the spring constant (k_R/k) for the mutated residue with the amino acid number shown. White corresponds to values of free energy predicted by the wild-type ENM. Red corresponds to an increase in $(G_{mut} - G_{wt})/|G_{wt}|$ (decreased value of G_{mut} comparing to G_{wt}), whereas blue corresponds to a decrease in ($(G_{mut} - G_{wt})/|G_{wt}|$) (increased value of G_{mut} comparing to G_{wt}).

B.1. Free energy mutation scan on apo structure. All active sites (besides res 163-167) appear on the mutation scan of the 6LU7 apo structure, a somewhat remarkable result considering that no ligand is present to emphasise the spacial nature of the active sites. They seem dynamically pre-disposed to dynamic allosteric communication, in agreement with the cross-correlation map (**Fig. 3**). Both termini display mutation peaks due to their spatial proximity to the active sites. Additionally, a very sharp peak is seen around res 187-192 where a free loop forming the active site is located. The six new regions seen on the cross-correlation map (res 17-32, 67-75, 77-91, 97-98, 120-131 and 201-214) form distinctive peaks as well on the mutation scan.

Furthermore, the experimentally identified residue 214 is signposted by these calcuations: its computational mutation generates the largest fluctuation free energy change: 0.29 % at $k_R/k = 4.00$ upon spring stiffening and -0.40 % when $k_R/k = 0.25$ upon spring stiffening - defining the amplitude of the colour bar. Although the catalytically paramount residue C145 is not as sharp as other peaks, it appears with greater strength in a higher mode summation (**SI, Fig. 9**). The map is mostly

qualitatively anti-symmetric around middle line (wild-type). However, the quantitative behaviour of three regions worthy of attention: the region of res 50, before 100, at res 150, as well as the sharpest region around res 187-192. Relaxation of stiffness at those points cause larger energy change than stiffening. A very narrow region, not identified on the cross-correlation map, at res 97-98, preceding a small beta sheet at res 101-104, appears sharply when local interaction strengths are relaxed. A new broad region of strong sensitivity to mutation appears on this map at res 261-29, which includes an alpha helix at res 261-274. This helix is located on the surface of the protein far away form the active site.

We also note 8 residues which are located on the homodimer chains' interface (**Tab. 1**), recalling that in the CAP homodimer residues located on the interface were critical in allosteric regulation (8). Especially responsive is res 14 located on the very first alpha helix.

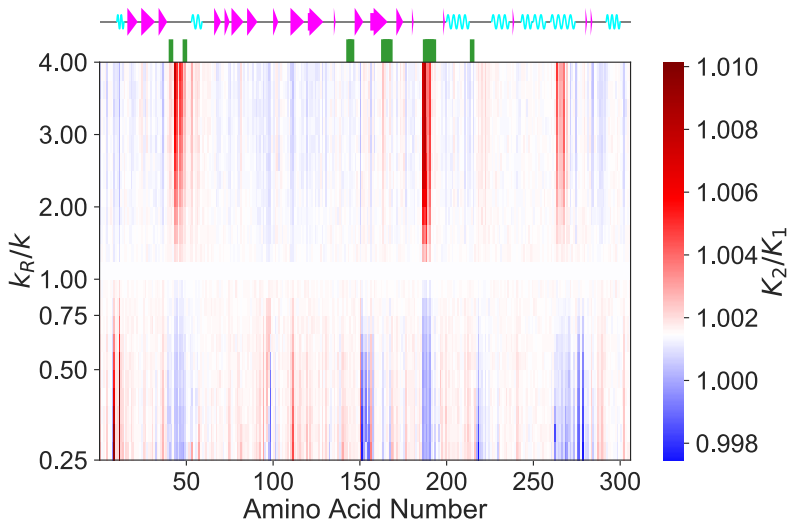

Fig. 5. A map for the global control space of allostery in M^{pro} calculated from the ENM over 25 modes. The map plots the change in cooperativity coefficient (K_2/K_1) due to the dimensionless change in the spring constant (k_R/k) for the mutated residue with the amino acid number shown. White corresponds to values of K_2/K_1 predicted by the wild-type ENM. Red corresponds to an increase in K_2/K_1 (stronger negative cooperativity), whereas blue corresponds to a decrease in K_2/K_1 (weaker negative cooperativity or positive cooperativity).

B.2. Allosteric free energy mutation scan. The first result from the ENM calculation of the allosteric free energy for binding-site occupation is that $K_2/K_1 \approx 1$ for the wild-type M^{pro} ENM. Therefore, this ENM model (over 25 softest modes) is non-cooperative. Nevertheless, we can identify regions that are sensitive to even slight change in local stiffness which again are around biologically active areas. All previously-marked active regions show-up to some extent; especially vivid is the region around catalytic residue H45 and, as already appeared in the apo mutation scan, the loop around the active site (res 187-192). **Residue N214 shows very weak allosteric control in this scan.** The local environment around residue 214 is mainly hydrophobic (**SI, Fig. 10**). Therefore, the experimentally-reported N214A mutation corresponds to a local structural stiffening (asparagine (N) is hydrophilic while alanine (A) is hydrophobic). This is indeed the region of parameter space ($k_R/k > 1$) where this mutation displays a weak effect in the ENM model, but a strong response upon relaxation.

C. 2-point mutational scans. It is of interest to explore the cooperative effect of two-point mutations in models of fluctuation allostery, as previous work has indicated that double mutations may combine non-linearly in control of the allosteric landscape of proteins (25). This numerical scan explores cases where mutations are made on only one of the single homodimeric chains: experimentally this is a possible, but not a trivial, task. However it can reveal contribution of each chain alone to fluctuation and allostery of the dimeric composite structure. The discussion of this sections refers to results presented in the 2-point scans of 6LU7 ENM in figure 6. In order to present the response to all double-mutations on a single 2D plot, the change in spring stiffness is not scanned, but replaced by a constant spring stiffening of 4.00, motivated both by a desire to model the effect of of small molecule/ligand binding and the agreement with the N214A mutation (**SI, Fig. 10**).

C.1. Free energy 2-point mutation scan on apo structure. The first measure, as in the single-point scans, is the difference in total free energy of the apo structure. As on the 1-point map for apo 6LU7 structure strong lines are observed (**Fig. 6a**) around the active residues with the additional regions around a small beta sheet (res 101-104), alpha helices (res 201-214 and 261-274) and loop region adjacent the latter helix. Note that, in the case of this protein, the 2-point mutations combine approximately linearly: the effect of the fist mutation (vertical lines) is not strongly affected by the second mutation. Once again, res 214 defines the highest value of 0.29 % as in the 1-point mutational scan of apo ENM (**Fig. 4**).

C.2. Allosteric free energy 2-point mutation scan. Finally, the effect of all double mutations on the allosteric free energy between the two active sites was calculated (**Fig. 6b**). A new strong control site that did not appear on the 2-point apo map (**Fig. 6a**) is found at the beginning of the second beta-sheet (res 35-39) on each chain. The apparent increased cooperativity for both mutations (in most cases) on chain A, and the same pattern of decreased cooperativity on chain B, is due to the formal symmetry breaking through choice of the first binding site at chain A. Mutation of residues around H45 are the exception, and have an opposite effect on allostery to all others. The dashed lines represent the 1-point allo scan in figure 5. In that scan K_2/K_1 values range from 0.998 to 1.010, while for this 2-point mutational scan the range is slightly increased. The same 2-point scan with $k_R/k = 0.25$ reached similar K_2/K_1 amplitude values (**SI, Fig. 11**).
This calculation draws attention to an additional advantage of the 2-point scans: while res 214 did not exhibit allosteric control on the 1-point map (**Fig. 5**), it appears as a strong line in all quadrants on the 2-point map. However, when the lines intersect in the cross-chain quadrant K_2/K_1 reaches almost wild-type value (corresponding to making exactly the same change on both monomers). This effect explains why we do not observe allosteric control of res 214 on the 1-point mutational map: evidently, when the bold lines intersect (identical mutation or binding on both domains) allosteric effects interfere destructively. Thus, in the 1-point allo scan for 6LU7 ENM amplitude of K_2/K_1 is lower than in the 2-point mutational allo scan.

3. Discussion: what does ENM tell us about SARS-CoV-2 Main Protease?

The ENM analysis reinforces previous findings in application to other proteins, that in the SARS-CoV-2 M^{pro} as well, local harmonic potentials within the equilibrium protein structure, but without mean structural change can identify already known biologically active sites. Furthermore, there is no need to have holo forms of the protein to locate those active sites, whose correlated dynamics are already clear in the apo form. Calculations of those sites where total free energies are sensitive to mutations converge well with the limit of the sum over normal modes. The convergence of calculations of control of the allosteric free energy itself is more subject to noise, being a difference-quantity, but

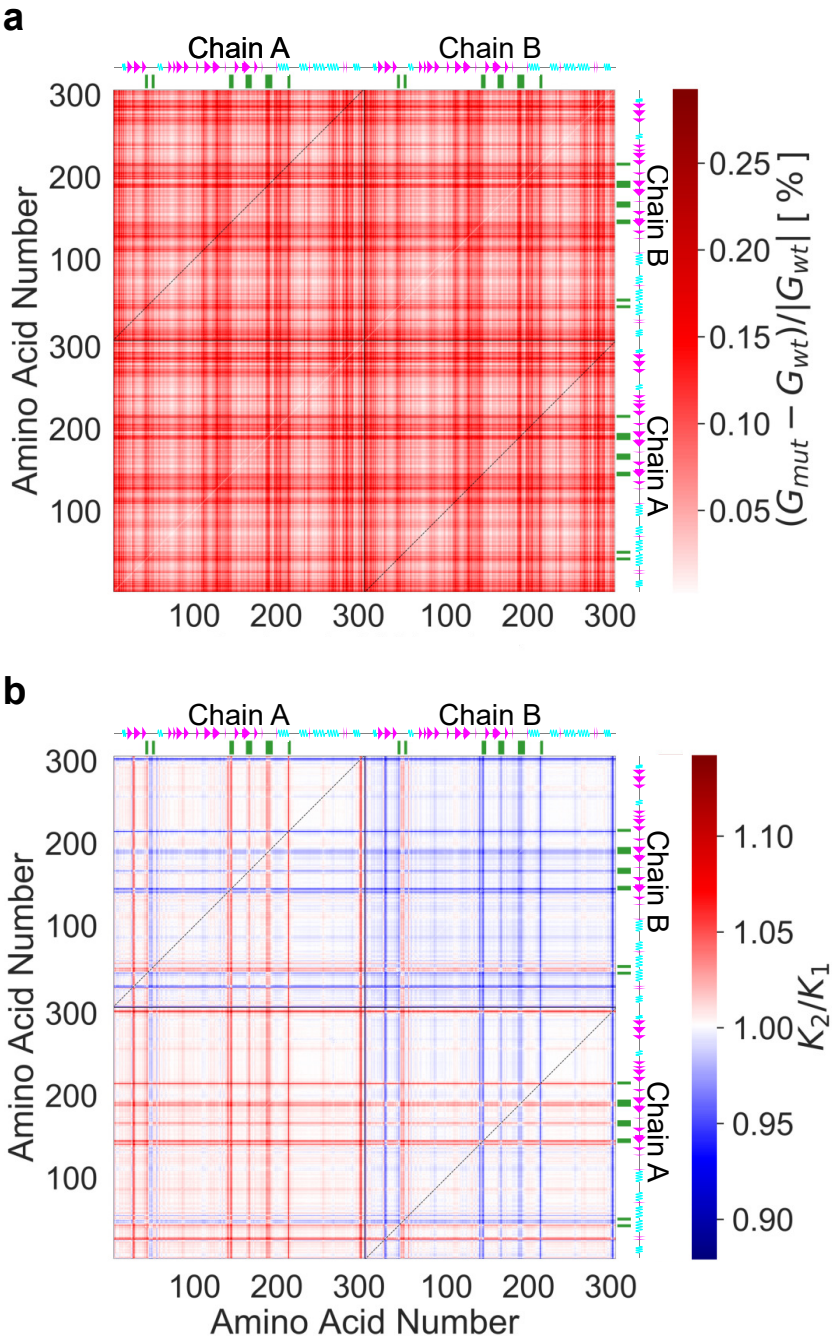

Fig. 6. 2-point mutational maps for 6LU7 ENM with all possible pairwise combinations of residue mutations with equal spring constant change $k_R/k = 4.00$ over the first real 25 modes.**a**, 2-point mutational map of 6LU7 apo ENM. All possible pairwise combinations of residue mutations with equal spring constant change. **b**, A map for the 2D global control space of allostery in M^{pro}. Black solid lines separate two homodimer chains, while dashed lines represent 1-point mutational scan results for the given spring constant change.

sufficiently to identify strong candidates for control regions (**SI, Fig. 9**).

The analysis shows that SARS-CoV-2 M^{pro} possesses a rich dynamical structure that supports several long-distance allosteric effects through thermal excitation of global normal modes. In particular the motions in the vicinity of two active sites are correlated within the first 25 non-trivial normal modes, especially in the singly-bound dimer. Although, at the level of ENM calculations, this does not lead

to cooperativity in the WT structure, it does render the protein susceptible to the introduction of cooperativity by mutation.

Our methodology is further supported by the ENM dynamics sensitivity to residue N214 mutation which has been previously experimentally verified to dynamically control SARS-CoV M^{pro}.

The ENM calculations have identified new sites whose local thermal dynamics dynamically correlate with those of the active sites, and which also appear on global maps for allosteric control by single or double mutations. The new candidate control regions are summarised in table 1. In particular, residues around the beta sheet (res 101-104) and the alpha helices(res 201-214 and 261-274) are novel, and distant from the active site (**SI, Fig. 12**). The position of these residues suggest them as possible candidates for non-competitive inhibiting binding sites. We also draw attention to eight residues located on M^{pro} interface surface (**Tab. 1**) as a potential dynamically allosteric control residues.

Computational studies such as this, therefore, accompany and support concurrent experimental programs of scanning for small-molecule binding candidates to the protein in question. We note that several candidate molecules, identified in a very recent large crystallographic fragment screen against SARS-CoV-2 M^{pro} (27), bind to regions suggested as dynamically sensitive control candidates in this study.

The ENM model employed was specific for the given inhibitor. Other ligands might, of course, show different behaviour in the corresponding holo structures and display other "hot-spots", however, the appearance of active regions, and their coupling, in the apo structure suggests that there are general properties that emerge from the global elastic structure of the protein. As well as providing specific information on the SARS-CoV-2 M^{pro} structure of the calculations reported there, the findings of this study also contribute to the large programme of research on fluctuation-induced allostery without conformational change. In particular the general question of the focusing of dynamic correlations between distant (so candidate allosteric) sites is solved in a highly specific way by this structure. It also constitutes a system for which double mutations contribute in a predominantly linear addition, in contrast to findings with other allosteric homodimers. This pattern includes the cancellation phenomenon we identified in the case of some single point mutations made identically and simultaneously in both monomers, whose cancellation in the 1-d scans can mask their potential sensitivity as target sites. Finally, the appearance of control regions on the exterior surface of proteins, with obvious pharmacological application, generates other general questions in the biophysics of fluctuation elasticity in globular proteins.

Materials and Methods

Normal Mode Analysis (NMA) of ENM describes protein motions around equilibrium and can be used to calculate the partition function for large scale harmonic thermal fluctuations in protein structure, including those responsible for allostery (28) . Two main approximations of NMA are:

- The structure fluctuates about at local energy minimum. Consequently no other structures beyond the given equilibrium can be explored.
- The force field everywhere arises from sums over ENM harmonic

The whole NMA method can be reduced to three steps:

1. Construct mass-weighted Hessian for a system. For a protein ENM the system consists of the co-ordinates of the C-alpha atoms (N) for each residue from the corresponding PDB structure.
2. Diagonalise the mass-weighted Hessian to find eigenvectors and eigenvalues of the normal modes.

3. Calculate the partition function (and so free energy) from the product over the normal mode harmonic oscillations.

The diagonalisation of the $3N \times 3N$ mass-weighted hessian matrix is written as

$$\mathbf{A}^{-1}\tilde{\mathbf{H}}\,\mathbf{A} = \mathbf{\Lambda}$$

where $\tilde{H}_{ij} = \dfrac{\partial^2 V_{ij}}{\partial r_i \sqrt{m_i} \partial r_j \sqrt{m_j}}$: the potential energy function V; distance between nodes r; node masses m. The eigenvectors of the mass-weighted Hessian matrix, columns of $\mathbf{A}$, are the normal mode eigenvectors $\mathbf{a}$.

$$\mathbf{A} = \begin{pmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_{3N} \\ | & | & & | \end{pmatrix}$$

$\mathbf{\Lambda}$ is a $3N \times 3N$ diagonal matrix with diagonal values equal to the associated normal modes' squared angular frequencies ω^2. The potential function used in this study is:

$$V_{ij} = \begin{cases} \dfrac{k_{ij}}{2}\left(r_{ij} - r_{ij}^{(0)}\right)^2 & r_{ij}^2 \leq r_c^2 \\ 0 & r_{ij}^2 > r_c^2 \end{cases}$$

where r_c is a cut-off radius, which for this work is set at 8 Å; while $r^{(0)}$ is the equilibrium distance between nodes derived form PDB crystallographic structure. For the wild-type protein, all spring constants are equal $k_{ij} = k = 1$ kcal Å^{-2} mol^{-1}.

Cross-correlation of Motion. The cross-correlation, C, is estimated between an ENM node pair as a normalised dot product sum between their normal mode eigenvectors over v modes.

$$C_{ij} = \sum_v \left(\frac{\mathbf{a}_i\,(v) \cdot \mathbf{a}_j\,(v)}{\sqrt{|\mathbf{a}_i\,(v)|^2\,|\mathbf{a}_j\,(v)|^2}} \right) \qquad [1]$$

C value of 1 implies perfectly correlated motion, -1 perfectly anti-correlated motion and 0 implies totally non-correlated motion.

Normal Mode Fluctuation Free Energy. Using statistical mechanics it is possible to calculate an estimate to the fluctuation free energy of a system using the frequency of vibrations such as the normal modes. For this method, the partition function fo quantum harmonic oscillator (29), Z, for normal mode k is given as

$$Z_k = \frac{\exp\left(-\frac{1}{2}\frac{\hbar\omega_k}{k_B T}\right)}{1 - \exp\left(-\frac{\hbar\omega_k}{k_B T}\right)}$$

where k_B is the Boltzmann's constant, $\hbar$ is the reduced Planck's constant, T is temperature in Kelvin and ω is, already mentioned, angular frequency. Gibbs free energy (for a given mode) expressed in terms of partition function, with an approximation of little change in volume, can be written as

$$G_k = -k_B T\, ln\left(\frac{1}{1 - \exp\left(-\frac{\hbar\omega_k}{k_B T}\right)}\right) + \frac{1}{2}\hbar\omega_k \qquad [2]$$

Ligand Dissociation Constant. When free energy change ΔG (**SI, Sec. C**) is known for a dissociation reaction, corresponding dissociation constant K can be estimated via

$$K = \exp\left(-\frac{\Delta G}{k_B T}\right) \qquad [3]$$

ACKNOWLEDGMENTS. ID is grateful for computational support from the University of York high performance computing service, VIKING. TCBM acknowledges support from the EPSRC(UK) through an Established Career Fellowship in the Physics of Life programme.

References

1. J Monod, J Wyman, JP Changeux, On the nature of allosteric transitions—a plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
2. A Cooper, DT Dryden, Allostery without conformational change. a plausible model. *Eur. Biophys. J.* **11**, 103–109 (1984).
3. TCB McLeish, TL Rodgers, MR Wilson, Allostery without conformation change: modelling protein dynamics at multiple scales. *Phys. biology* **10**, 056004 (2013).
4. C Schaefer, AC vonderHeydt, T McLeish, The 'allosteron' model for entropic allostery of self-assembly. *Philos. Transactions* **373**, 20170186 (2018).
5. N Go, T Noguti, T Nishikawa, Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci.* **80**, 3696–3700 (1983).
6. B Brooks, M Karplus, Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc. Natl. Acad. Sci.* **82**, 4995–4999 (1985).
7. I Bahar, A Atilgan, B Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* **2**, 173–181 (1977).
8. TL Rodgers, et al., Modulation of global low-frequency motions underlies allosteric regulation: demonstration in crp/fnr family transcription factors. *PLoS biology* **11** (2013).
9. TL Rodgers, et al., Ddpt: a comprehensive toolbox for the analysis of protein motion. *BMC Bioinforma.* **14**, 183 (2013).
10. L Zhang, et al., Crystal structure of sars-cov-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* (2020).
11. T Muramatsu, et al., Sars-cov 3cl protease cleaves its c-terminal autoprocessing site by novel subsite cooperativity. *Proc. Natl. Acad. Sci.* **113**, 12997–13002 (2016).
12. T Pillaiyar, M Manickam, V Namasivayam, Y Hayashi, SH Jung, An overview of severe acute respiratory syndrome–coronavirus (sars-cov) 3cl protease inhibitors: Peptidomimetics and small molecule chemotherapy. *J. medicinal chemistry* **59**, 6595–6628 (2016).
13. Z Jin, et al., Structure of m^{pro} from covid-19 virus and discovery of its inhibitors. *Nature* (2020).
14. W Dai, et al., Structure-based design of antiviral drug candidates targeting the sars-cov-2 main protease. *Science* (2020).
15. Z Jin, et al., Structural basis for the inhibition of sars-cov-2 main protease by antineoplastic drug carmofur. *Nat. Struct. & Mol. Biol.*, 1–4 (2020).
16. L Di Paola, A Giuliani, Mapping active allosteric loci sars-cov spike proteins by means of protein contact networks. *arXiv preprint arXiv:2003.05200* (2020).
17. YW Chen, CPB Yiu, KY Wong, Prediction of the sars-cov-2 (2019-ncov) 3c-like protease (3cl pro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. *F1000Research* **9** (2020).
18. X Ou, et al., Characterization of spike glycoprotein of sars-cov-2 on virus entry and its immune cross-reactivity with sars-cov. *Nat. communications* **11**, 1–12 (2020).
19. AT Ton, F Gentile, M Hsing, F Ban, A Cherkasov, Rapid identification of potential inhibitors of sars-cov-2 main protease by deep docking of 1.3 billion compounds. *Mol. Informatics* (2020).
20. M Kandeel, M Al-Nazawi, Virtual screening and repurposing of fda approved drugs against covid-19 main protease. *Life Sci.*, 117627 (2020).
21. E Estrada, Topological analysis of sars cov-2 main protease. *bioRxiv* (2020).
22. M Hofmarcher, et al., Large-scale ligand-based virtual screening for sars-cov-2 inhibitors using deep neural networks. *Available at SSRN 3561442* (2020).
23. AD Elmezayen, A Al-Obaidi, AT Şahin, K Yelekçi, Drug repurposing for coronavirus (covid-19): in silico screening of known drugs against coronavirus 3cl hydrolase and protease enzymes. *J. Biomol. Struct. Dyn.*, 1–12 (2020).
24. J Shi, et al., Dynamically-driven inactivation of the catalytic machinery of the sars 3c-like protease by the n214a mutation on the extra domain. *PLoS computational biology* **7** (2011).
25. PD Townsend, et al., The role of protein-ligand contacts in allosteric regulation of the escherichia coli catabolite activator protein. *J. Biol. Chem.* **290**, 22225–22235 (2015).
26. PD Townsend, et al., Global low-frequency motions in protein allostery: Cap as a model system. *Biophys. reviews* **7**, 175–182 (2015).
27. Diamond Synchrotron, Main Protease Structure and XChem Fragment Screen (Link) (2020)

Online: accessed 20/05/2020.
28. I Bahar, TR Lezon, A Bakan, IH Shrivastava, Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem. reviews* **110**, 1463–1497 (2010).
29. SJ Blundell, KM Blundell, *Concepts in thermal physics*. (OUP Oxford), (2009).
30. JA McCammon, BR Gelin, M Karplus, PG WOLYNES, The hinge-bending mode in lysozyme. *Nature* **262**, 325–326 (1976).
31. A Nicolaï, P Delarue, P Senet, Theoretical insights into sub-terahertz acoustic vibrations of proteins measured in single-molecule experiments. *The journal physical chemistry letters* **7**, 5128–5136 (2016).
32. D Eisenberg, E Schwarz, M Komarony, R Wall, Amino acid scale: Normalized consensus hydrophobicity scale. *J Mol Biol* **179**, 125–142 (1984).

Supplementary Information

A. PDB structures. Atoms were manually removed from the original 6LU7 PDB file to make up the three protein forms: ligand-free (apo); singly-bound ligand to chain A (holo1); each chain, A and B, posses one ligand (holo2).

B. DDPT. This section states the routine we used to generate and analyse M^{pro} (6LU7) ENM via Durham Dynamic Protein Toolbox (DDPT) (9). As an example, the routine is carried for 6LU7 holo1 form.
First, we generate ENM interaction matrix using GENENMM function in DDPT:

GENENMM -pdb 6lu7_holo1.pdb -c 8 -het -fcust ffile.txt -mass -ca -res

The crystallographic structure of a protein is assigned by `-pdb` flag. Cut-off distance (`-c`, by default 12 Å) was set to 8 Å and heteroatoms (`HETATOM` in PDB notation) from the PDB file were included using `-het`, if required. Hookean spring strengths (by default 1 kcal mol $^{-1}$ Å^{-2}) around selected residues can be altered via `-fcust` flag. The associated `ffile.txt` specifies which residue springs' moduli are modified. By default, all ENM nodes' masses (atoms from PDB structure are converted into ENM nodes) are set to the same fixed value, 1 amu, but actual atomic masses (`-mass`) were used in our ENM. `-ca` flag creates ENM model where amino acids' C_α atoms form nodes. In combination with `-ca`, `-res` flag assigns the C_α nodes the whole residue mass. GENENMM Generates interaction matrix for the ENM written into `matrix.sdijf` file.
Next, DIAGSTD function diagonalises the interaction matrix.

DIAGSTD -i matrix.sdijf

where `-i` flag specifies input. The product of the diagonalisation is written into `matrix.eigenfacs` file.
We calculate inter-atom distance and cross-correlation of motion map for the C_α ENM of the protein using SPACING and CROSCOR functions, respectively.

SPACING -pdb 6lu7_holo1.pdb -ca -het

and

CROSCOR -i matrix.eigenfacs -s 7 -e 31 -het

where `-s` and `-e` flags indicate the first and the last normal modes to include in calculation. Note, the first six normal mode frequencies, which correspond to translational and rotational motion, are equal to zero.
Finally, fluctuation free energies for the set of the normal modes were calculated via FREQEN as follows

FREQEN -i matrix.eigenfacs -s 7 -e 106

Temperature (T) value used to calculate the partition function Z and Gibbs free energy G was 298 K, the default value for FREQEN function.

B.1. Fluctuation Free Energy Approximation. For convenience, DDPT estimates fluctuation free energy for each normal mode in dimensionless units $\frac{G}{k_B T}$ as follows

$$\frac{G}{k_B T} = -ln\left(\frac{1}{1-\exp\left(-\frac{\hbar\omega}{k_B T}\right)}\right) + \frac{1}{2}\frac{\hbar\omega}{k_B T}$$

Normal mode frequencies in proteins are in acoustic regime (less than 10 THz) (30, 31). This fact is especially true for the slowest modes we investigated in this study. In the classical limit, which we are in, $\frac{\hbar\omega}{k_B T} \ll 1$ at 298 K temperature. Therefore, FREQEN function in DDPT employs the following expression for fluctuation free energy calculation

$$\frac{G}{k_B T} \approx -ln\left(\frac{1}{1-\exp\left(-\frac{\hbar\omega}{k_B T}\right)}\right) \qquad [4]$$

Moreover, when we estimate difference (ΔG) and difference of a difference ($\Delta\Delta G$) in fluctuation free energy (**Eq. 5**) significance of $\frac{1}{2}\frac{\hbar\omega}{k_B T}$ term diminishes drastically.

C. Fluctuation Energy Convergence. Allosteric free energy change is calculated using fluctuation free energy (**Eq. 6**) change for three forms of the protein:

$$\begin{aligned}\Delta\Delta G &= \Delta G_2 - \Delta G_1 \\ &= (G_{holo2} - G_{holo1}) - (G_{holo1} - G_{apo}) \\ &= G_{holo2} - 2G_{holo1} + G_{apo}\end{aligned} \quad [5]$$

For apo form, fluctuation free energy curve is smooth and stable (**Fig. 7**). However, in figure 8 we saw instability in dissociation constant ratio due-to the difference of a difference in the fluctuation free energies. K_2/K_1 values between 25 and 35 modes fluctuate stably in 1.00-1.02 range. For this reason and three other stated in section 2 of the paper, we have chosen to stop summing at the 25 real modes.
1-point mutational scan for apo form is stable in respect the biologically active sites for increasing value of modes ((**Fig. 9a**)). The same mutational scan but for allosteric energy change is less stable (**Fig. 9b**). Nevertheless, the bioactive sites persist and we must take into account the sensitivity of K_2/K_1 already mentioned in figure 8. 2-point mutational scan for allosteric control is stable for different k_R/k values over slowest 25 modes (**Fig. 6, 11**). As stiffening response is change, we observed sign change on in the allosteric response but active residue pattern stayed the same.

D. SARS-CoV-2 M^{pro} Analysis. The choice of the spring constant moduli change for the 2-point mutational scans was based on N214A mutation in SARS-CoV M^{pro} that dynamically inactivated the protein (24). We visually investigated hydrophobic environment around residue 214 to choose how k_R/k in our ENM corresponds to this experimental mutation (**Fig. 10**).
The identified candidate regions that exhibit allosteric activity in our ENM study (**Tab. 1**) were visualised in 3D space (**Fig. 12**).

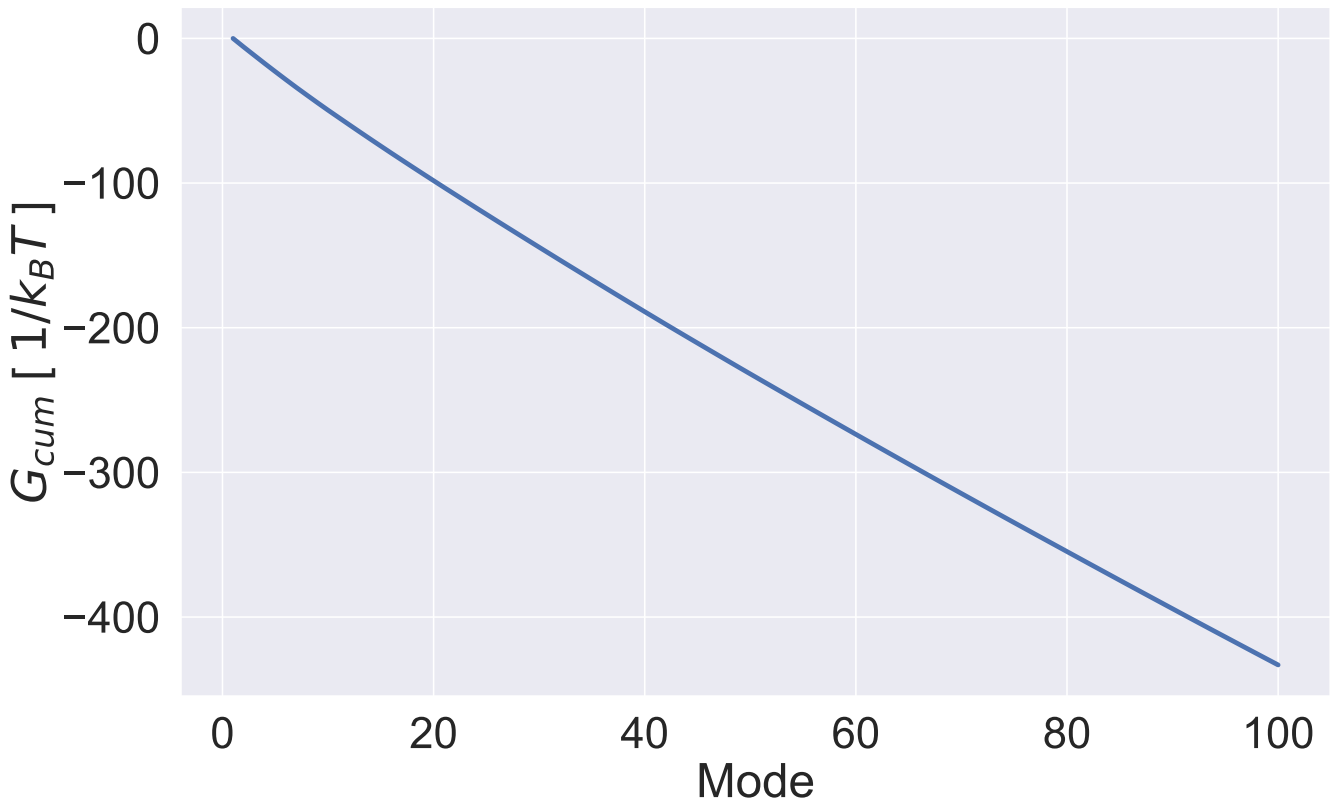

Fig. 7. Wild-type ligand-free SARS-CoV-2 M^{pro} ENM fluctuation free energy dependence on mode summation at 298 K. All ENM C_α node masses correspond to amino acid mass, i.e. whole residue mass. Cut-off distance is 8 Å. All ENM spring constants are equal 1 kcal mol^{-1} $Å^{-2}$.

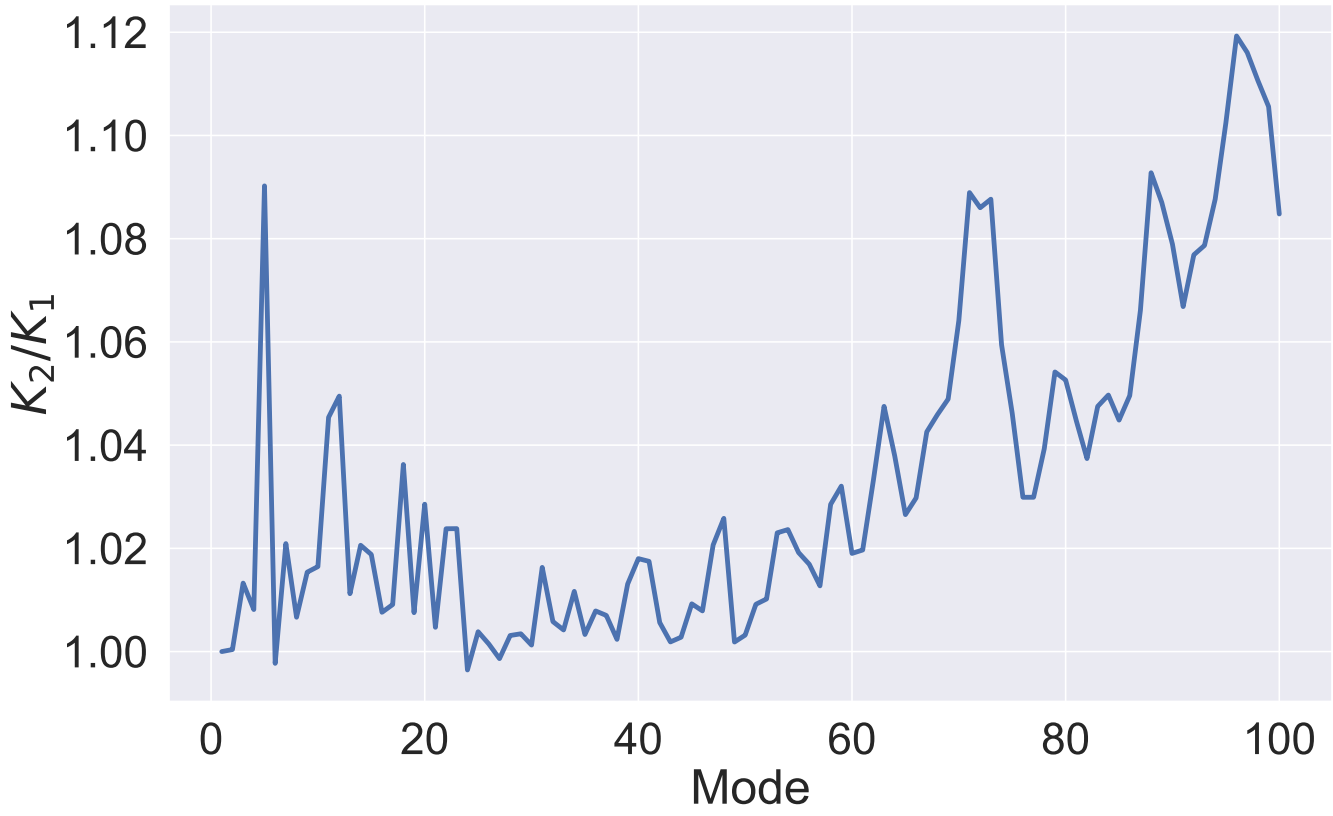

Fig. 8. Wild-type SARS-CoV-2 M^{pro} ENM cooperativity dependence on mode summation. All ENM node masses correspond to amino acid mass a node been derived from; while ligand nodes mass is assigned based on the element. Cut-off distance is 8 Å. All ENM spring constants are equal 1 kcal mol^{-1} $Å^{-2}$.

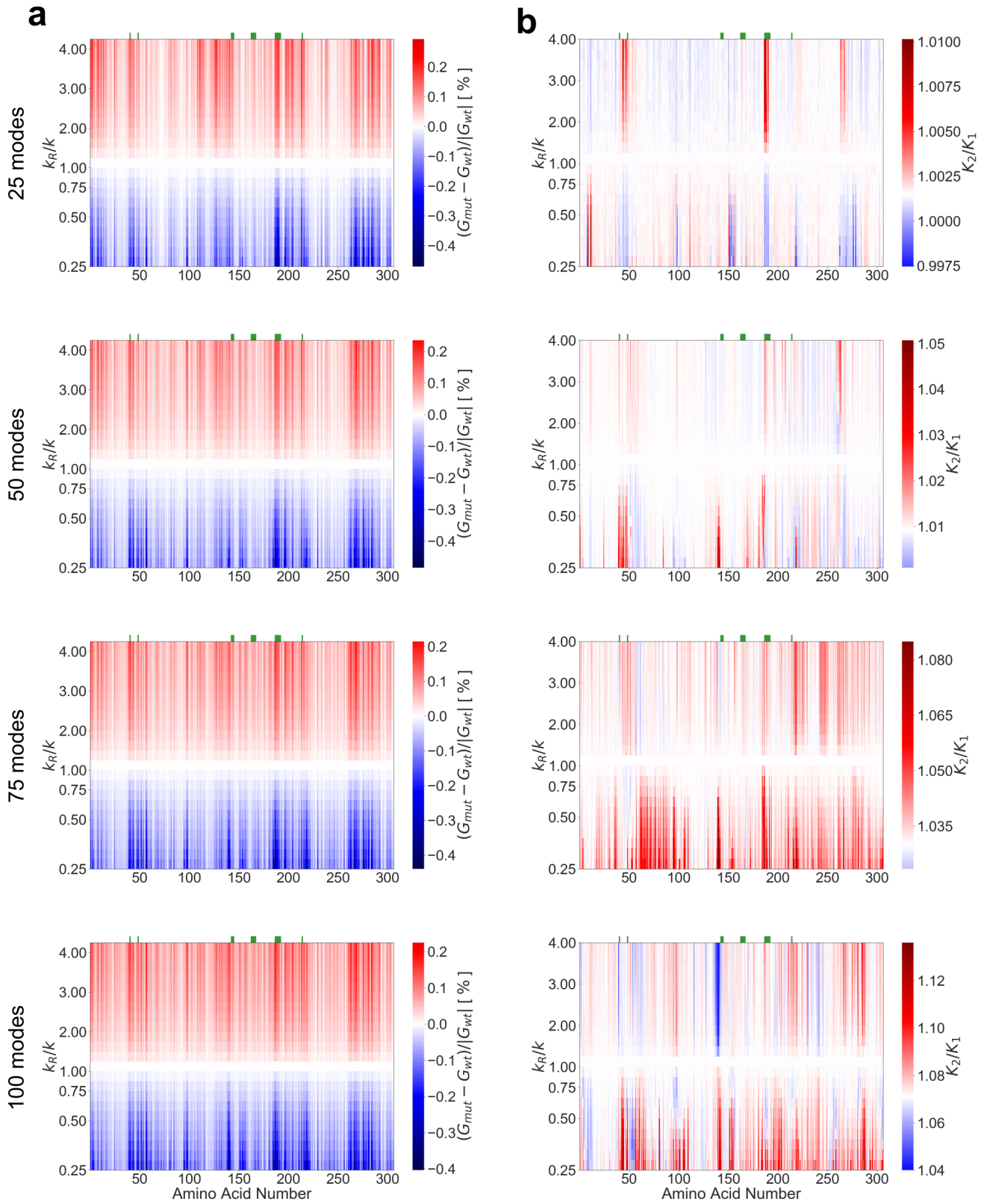

Fig. 9. 1-point mutational scan variation in M^{pro} (6LU7) ENM with the real mode summation. **a**, 1-point mutational scan of apo ENM over 4 increasing cumulative mode values. **b**, A map for the global control space of allostery in M^{pro} calculated from the ENM over 4 increasing cumulative mode values.

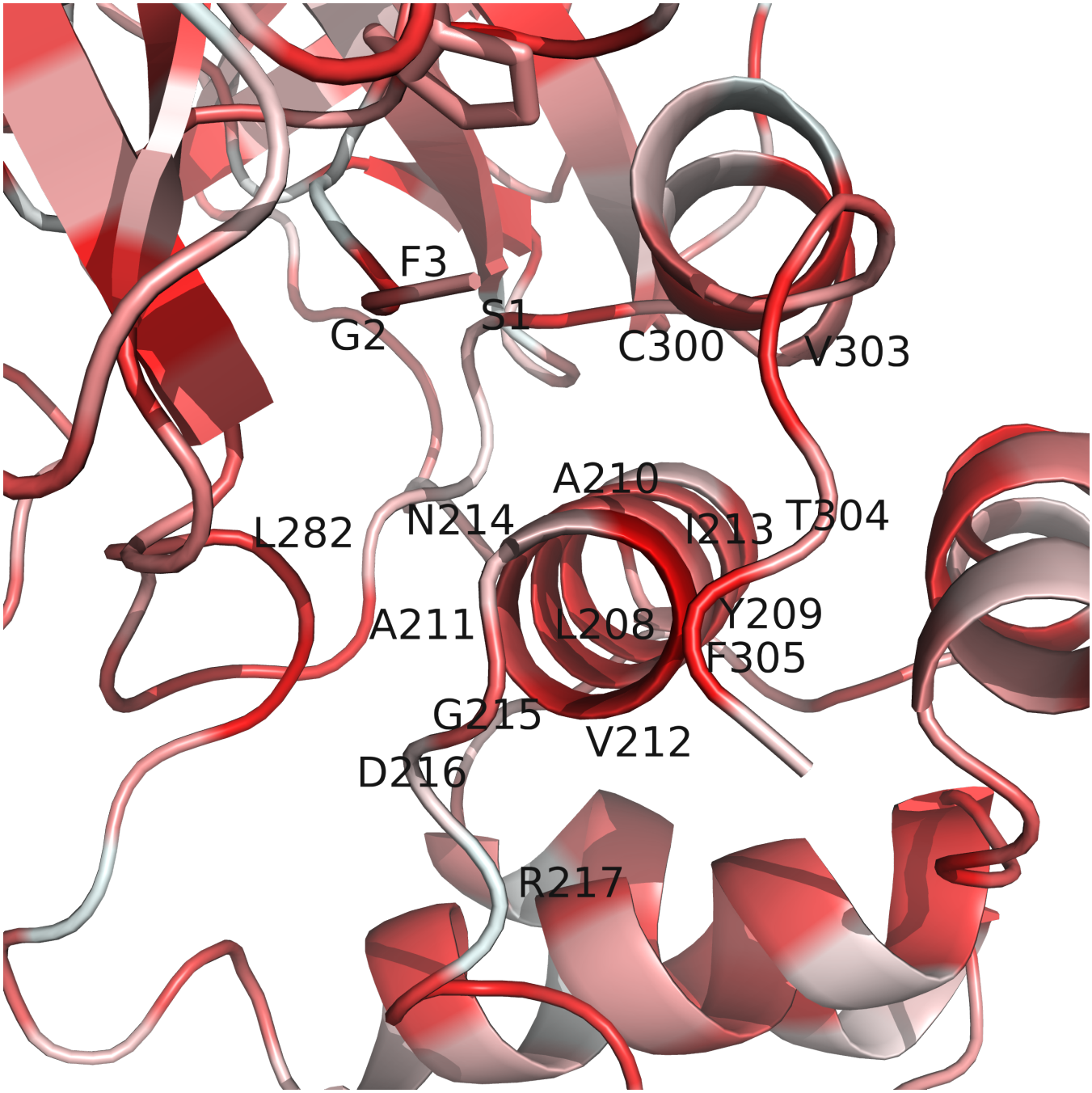

Fig. 10. Hydrophobic environment around residue 214 in SARS-CoV-2 M^{pro} based on Eisenberg amino acid hydrophobicity scale (32). Red illustrates hydrophobic residues while white shows less hydrophobic (hydrophilic) residues. The residues within 8 Å radius are labelled with the single amino acid letter code and residue number.

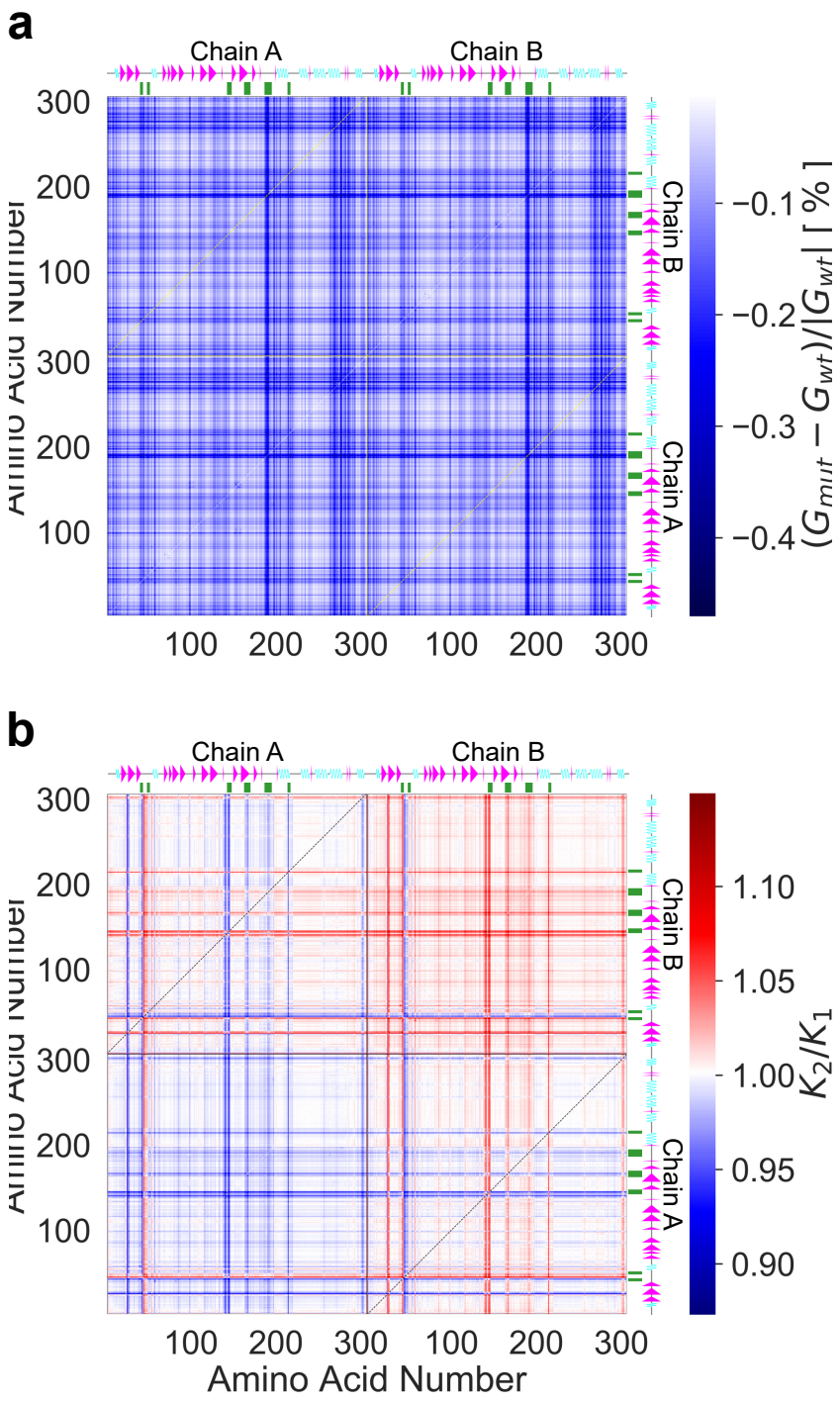

Fig. 11. 2-point mutational maps for 6LU7 ENM with all possible pairwise combinations of residue mutations with equal spring constant change $k_R/k = 0.25$ over the first real 25 modes. **a**, 2D scan for M^{pro} apo ENM. Yellow solid lines separate two homodimer chains, while dashed lines represent 1-point mutational scan results for the given spring constant change. **b**, A map for the 2D global control space of allostery in M^{pro} calculated from the ENM. Black solid lines separate two homodimer chains, while dashed lines represent 1-point mutational scan results for the given spring constant change. The secondary structure of M^{pro} indicated along the residue axes, with cyan waves indicating alpha helices, and magenta triangles indicating beta sheets. The green ticks on the axis indicate the location of biologically active residues (**Tab. 1**).

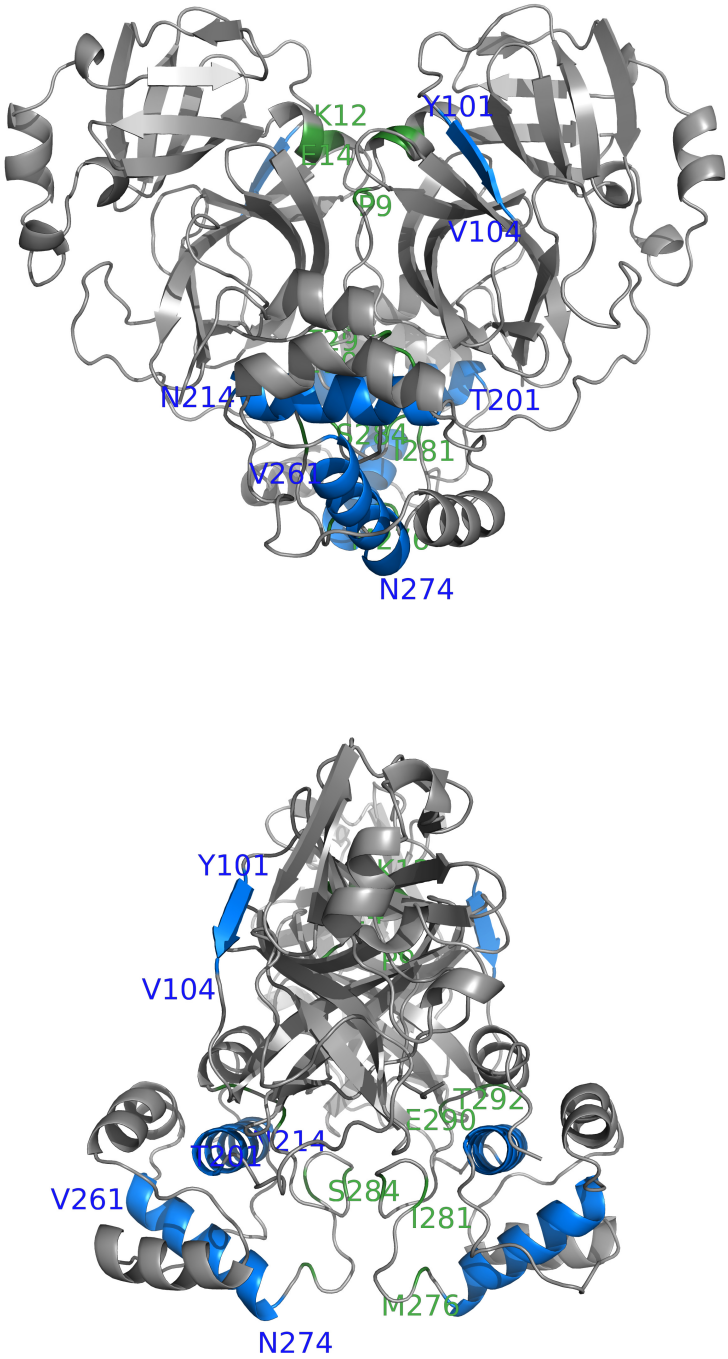

Fig. 12. Identified active regions in M^{pro}, (*top*) front view and (*bottom*) side, that are distant from the biologically active sites using 6LU7 ENM. The secondary structures around the identified regions distant form the active sites are coloured in blue. Start and end residues of the coloured secondary structures are labelled in single amino acid code with residue number on chain A. The identified active residues at the chains' interface are coloured green and labelled in single amino acid code with residue number on chain B. The active region residues can be found in table 1 of the manuscript.

E. Simulation Data. Numerical results used to construct figures 3, 4, 5 and 6 are attached to this manuscript in Excel spreadsheet form.

SI Dataset S1 (**6lu7_croscor-dist_apo_25modes.xlsx**)
Figure 3 left: Cross-correlation of motion and residue-reside distance map for 6LU7 apo ENM over 25 fluctuation modes.

SI Dataset S2 (**6lu7_croscor-dist_holo1_25modes.xlsx**)
Figure 3 middle: Cross-correlation of motion and residue-reside distance map for 6LU7 holo1 ENM over 25 fluctuation modes..

SI Dataset S3 (**6lu7_croscor-dist_holo2_25modes.xlsx**)
Figure 3 right: Cross-correlation of motion and residue-reside distance map for 6LU7 holo2 ENM over 25 fluctuation modes..

SI Dataset S4 (**6lu7_1D-apo+allo_25modes.xlsx**)
Figure 4 and figure 5: 1-point mutational scan results for 6LU7 apo and allosteric maps over 25 fluctuation modes.

SI Dataset S5 (**6lu7_2D-apo+allo_25modes.xlsx**)
Figure 6a-b: 2-point mutational scan results for 6LU7 apo and allosteric maps over 25 fluctuation modes.