Dynamical Cooperativity of Ligand-Residue Interactions Evaluated with the Fragment Molecular Orbital Method

Shigenori Tanaka[1*], Shusuke Tokutomi[1], Ryo Hatada[2], Koji Okuwaki[2], Kazuki Akisawa[2], Kaori Fukuzawa[3,4,5], Yuto Komeiji[6], Yoshio Okiyama[7], Yuji Mochizuki[2,5]

[1] *Graduate School of System Informatics, Department of Computational Science, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe 657-8501, Japan*

[2] *Department of Chemistry and Research Center for Smart Molecules, Faculty of Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan*

[3] *School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa-Ku, Tokyo 142-8501, Japan*

[4] *Department of Biomolecular Engineering, Graduate School of Engineering, Tohoku University, 6-6-11 Aoba, Aramaki, Aoba-ku, Sendai 980-8579, Japan*

[5] *Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan*

[6] *Biomedical Research Institute, AIST, Tsukuba Central 6, Tsukuba, Ibaraki 305-8566, Japan*

[7] *Division of Medicinal Safety Science, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa 201-9501, Japan*

E-mail: tanaka2@kobe-u.ac.jp (Shigenori Tanaka); fullmoon@rikkyo.ac.jp (Yuji Mochizuki)

Abstract

By the splendid advance in computation power realized with Fugaku supercomputer, it has become possible to perform *ab initio* fragment molecular orbital (FMO) calculations for thousands of dynamical structures of a protein-ligand complex in a parallelized way. We have thus carried out the electron-correlated FMO calculations for a complex of the 3C-like (3CL) main protease (M^{pro}) of the new coronavirus (SARS-CoV-2) and its inhibitor N3 incorporating the structural fluctuations sampled by classical molecular dynamics (MD) simulation in hydrated condition. Along with a statistical evaluation of inter-fragment interaction energies (IFIEs) between the N3 ligand and surrounding amino-acid residues for a thousand of dynamical structure samples, we have applied in this study a novel approach based on the principal component analysis (PCA) and the singular value decomposition (SVD) to the analysis of IFIE data in order to extract the dynamically cooperative interactions between the ligand and residues. We have found that the relative importance of each residue is modified via the structural fluctuations and that the ligand is bound in the pharmacophore in a dynamical manner through collective interactions formed by multiple residues, thus providing a new insight into structure-based drug discovery.

INTRODUCTION

Computational analysis of protein-ligand interactions is a key issue in structure-based, rational drug discovery.[1] Through long history of this direction of scientific research, a variety of theoretical approaches based on molecular docking,[2-4] molecular dynamics[5-7] and quantum chemical calculations[1] have been developed. In these approaches, physicochemical properties such as scoring function,[8] binding free energy,[9] ligand-residue interaction and interaction fingerprint[10] have been employed to quantitatively represent the affinity between protein and ligand molecules. The significance of these computational measures is extensively recognized also in the urgent drug discovery issues concerning the COVID-19 pandemic.[11]

The COVID-19 disease caused by the new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is still the worldwide critical issue, as reported by the World Health Organization (WHO).[12] A great amount of works have been conducted in various fields against this difficult situation. In the pharmaceutical approaches, several proteins such as the 3C-like (3CL) main protease (M^{pro}) of SARS-Cov-2 have attracted considerable attention because they are potential targets of antiviral drugs, and hence a number of computational studies have been performed to find effective inhibitors against them.[11] For example, at the early stage of COVID-19 outbreak, Xu et al. [13] showed that nelfinavir, which was known as an antiretroviral drug for the treatment of the human immunodeficiency virus (HIV), may be a promising inhibitor to M^{pro}, based on the homology modeling of SARS-CoV-2 M^{pro}

and the docking simulations of thousands of small molecule drugs. Jin et al.[14] identified a new peptide-like inhibitor called N3 on the basis of structure-assisted and computer-aided drug design strategy, and then determined the crystal structure of M^{pro}-N3 complex (registered as 6LU7 in Protein Data Bank (PDB)) in which the N3 ligand is covalently bonded to Cys145 via reactive Michael addition. Following this availability of 6LU7 structure, a number of computational studies primarily including molecular dynamics (MD) simulations were carried out[15-24] towards the elucidation of ligand binding mechanism and discovery of effective inhibitors. In the quantum chemical studies, we[24] performed *ab initio* fragment molecular orbital (FMO)[25-29] method-based analyses of interactions between the N3 inhibitor and amino-acid residues in the pharmacophore, in which pair interaction energy (PIE)[27] or inter-fragment interaction energy (IFIE)[28,29] was employed to identify crucial residues having site-specific interactions with N3 through, e.g., hydrogen bond.

While we additionally performed detailed PIE decomposition analyses (PIEDA[30,31]) including also the implicit hydration effect by means of the Poisson-Boltzmann (PB) model[32,33] in Ref. 24, the structural fluctuation effect (which should be essential in physiological condition) was not incorporated there because we performed only a single-structure FMO calculation with the 6LU7 crystal structure data.[14] Therefore, we[34] subsequently carried out the same FMO approach for the M^{pro}-N3 complex again, but using a thousand of structure samples generated by classical MD

simulation. This demanding task was firstly realized through large capacity computing with the novel

use of a massively parallel computing resource of Fugaku supercomputer[35] at RIKEN in Kobe.

Now, we are in the new era for FMO application to protein-ligand binding analysis. We can perform

the FMO-IFIE analysis on thousands of snapshot structures of protein-ligand complex system

generated by MD simulation. Our present problem is what we can do using these dynamical interaction

data with quantum chemical accuracy. In addition to our earlier study[34] mainly concerning the

dynamical average and fluctuation of protein-ligand IFIEs, we here employ two statistical techniques

of data analysis, principal component analysis (PCA) and singular value decomposition (SVD),[36-38]

to elicit the collective dynamics of interactions between a ligand molecule and surrounding amino-

acid residues. Through these novel approaches, we anticipate that cooperative interactions between

the ligand and multiple residues can be described dynamically to represent the specific binding modes

of ligand molecule in the pharmacophore. These kinds of computational descriptions are recently

attracting much attention; for example, a ligand molecule is often bound by multiple interaction

points[20,21], thus indicating the importance of molecular interaction fingerprint[10] and its dynamical

extension. Also including the possibility of describing the dynamical interaction network as observed

in hydrogen-bonding systems,[39] the present quantum-chemical approach can thus pave a new avenue

for computational drug discovery. The computational protocols in the present study are detailed in the

following, along with some new findings concerning the binding mechanism between M^{pro} and its possible inhibitors.

MATERIALS AND METHODS

Preparation of structures of M^{pro}-N3 complex.

We employed the crystal structure of the M^{pro}-N3 complex (PDB entry: 6LU7), from which we extracted a monomer unit, and processed it in a standard manner with the MOE software;[40] that is, we added missing hydrogens and optimized them with the AMBER10:EHT[41] force field (refer also to Ref. 24 for details), where we assigned a doubly-protonated cationic model (HIP) to His163. This structure (see Fig. 1) is referred to as "static" one hereafter. A thousand of "dynamic" structures were then prepared by the MD simulation method with the AMBER16 program[42] (with the AMBER10:EHT force field), where counter ions (Na$^+$) were generated to electrically neutralize the system. The durations of MD run were 50 ps for the thermal elevation from 0 K to 310 K (NVT ensemble), 50 ps for the density relaxation (NPT ensemble), 1 ns for the equilibration (NPT ensemble), and 100 ns for the production process (at 310K and 1 atm with NPT ensemble). From trajectories of the production run, a thousand snapshots were sampled with the interval of 100 ps. This fairly long-time MD sampling could cover larger configurational spaces and thus provides dependable statistical evaluations of dynamically fluctuating interaction energies between ligand and amino-acid residues.

All the MD simulations were performed on the TSUBAME3.0 GPU-accelerated computer at Tokyo

Institute of Technology. Each sample structure was shaped to a droplet form with water molecules

within 4 Å of M^{pro} for the FMO calculations, where all the counter ions were retained to keep the

system electrically neutral; this criterion of water layer thickness was adopted according to the

literature.[43-45]

Fragment molecular orbital (FMO) method.

The electronic energy of two-body FMO expansion[25-29] can be expressed as

$$E_{FMO} = \sum_I E_I' + \sum_{I>J} \Delta\tilde{E}_{IJ} \qquad (1)$$

where *I* and *J* are fragment indices that represent amino-acid residues or (parts of) ligand molecule in

the present study. The first term of eq. 1 refers to the renormalized fragment monomer energy, and the

second term corresponds to the fragment interaction energy termed as PIE[27] or IFIE.[28,29] The electron

correlation effect represented by the dispersion (DI) energy is taken into account at the second-order

Møller-Plesset perturbation (MP2) level[46-49] in this study. We adopted the partial renormalization

(PR)[50] to reduce the tendency of overestimation in MP2-based DI energy.[24] As shown in Fig. S1 in

Supporting Information (SI), the N3 ligand was segmented to 5 fragments[24] and has a covalent bond

to Cys145 (C145; hereafter, 1-letter code for amino-acid residues is primarily used for the simplicity

of representation) in the 6LU7 structure.[14] The numbers of fragments were 311 for the M^{pro}-N3

complex and about 1400 for the water layer, respectively. The basis set used was 6-31G*[51], a standard

choice in recent FMO studies. A series of FMO-MP2/6-31G* calculations with the PIEDA option were

performed with the ABINIT-MP program[28,29] on the supercomputer Fugaku.[35] Typical job timing for a single structure sample was 0.6 h with 192 nodes (or a half rack) of Fugaku, and the concurrent job processing led to only 6 h for the completion of all the 1000 structure samples. In addition to the structures with the covalent bond between the N3 ligand and C145, we have also carried out the FMO calculations for the structures without the covalent bond[24] in order to investigate the case when the ligand molecule is bound with non-covalent interactions.

It is noted that an amino-acid "residue" corresponds to one "fragment" in this study. When we make the interpretation of IFIE for the ligand-residue interaction, caution is in order for the fact that the fragmentation of usual FMO calculation for protein is made at the single bond between the carbonyl carbon and the alpha carbon (with side chain)[52] and thus that a misleading assignment shift of interacting fragment may occur when the carbonyl oxygen forms a hydrogen bond. This issue was carefully handled as was done in Ref. 24. We obtained the averages and standard deviations of IFIE values calculated for the "dynamic" structures, where the ligand-residue IFIE was obtained through the summation of the contributions from the five fragments composing the N3 ligand (see Fig. S1 in SI).

Statistical analyses of dynamical FMO-IFIE data.

Through the FMO calculations for the protein-ligand complex whose structure is temporally varied, we obtain the time-evolving IFIE matrix $X = (x_{ij})$, where $i = 1,...,m$ refers to the time step and $j = $

1,…,n represents each residue in the protein, when we consider the ligand-residue interactions; here, the summation over the IFIEs for the five fragments of N3 ligand has been taken. The corresponding variance-covariance matrix $Z = (z_{pq})$ ($1 \leqq p,\ q \leqq n$=306 in the case of the presence of covalent bond to C145) is then calculated by

$$z_{pq} = \frac{1}{m} \Sigma_i (x_{ip} - \overline{x_p})(x_{iq} - \overline{x_q}) \qquad (2)$$

where $\overline{x_p}$ denotes the temporal average of x_{ip} over *m*=1000 step.

In the principal component analysis (PCA), the variance-covariance matrix is decomposed as

$$Z = A\Lambda A^T \qquad (3)$$

as in Fig. 2, where Λ is a diagonal matrix whose elements λ_i represent the eigenvalues of *i*-th components (PC*i*) and A is a unitary matrix composed of the set of eigenvectors (superscript T denotes the transpose). The weight of the *j*-th residue can then be defined by

$$C_j = \Sigma_i \lambda_i a_{ij}^2 \qquad (4)$$

through summation over the eigenstates *i* using the components a_{ij} in A.

On the other hand, in the singular value decomposition (SVD),[36-38] the time-evolving IFIE matrix X is decomposed as (see Fig. 2)

$$X = U\Sigma V^T \qquad (5)$$

where Σ contains the singular values σ_i for the diagonal elements; U and V refer to matrices composed of the left and right singular vectors, respectively. In the present case, U and V represent

the characteristics concerning the time and residue variations, respectively. The i-th components

associated with the singular values σ_i are designated as SVi in the following. The weight of the j-th

residue can then be defined by

$$D_j = \sum_i \sigma_i v_{ij}^2 \qquad (6)$$

through summation over i using the components v_{ij} in V.

RESULTS AND DISCUSSION

Structural dynamics and temporal variations of IFIEs.

Figure 3(a) illustrates a superimposed view of 100 ns fluctuating structures of the N3 ligand and

surrounding residues sampled with 10 ns intervals in the presence of the covalent bond between N3

and M^{pro}. Because of the covalent bond to C145 of M^{pro}, fluctuations of Fragments 3 and 4 of the

ligand (see Fig. S1 in SI) are suppressed, whereas those of the terminal Fragments 1 and 5 are relatively

larger; refer also to Fig. S2 in SI for the root mean square deviation (RSMD) of trajectories.

On the other hand, Figure 3(b) illustrates the structural changes of the N3 ligand in the absence of

the covalent bond to C145. Naturally, the ligand showed larger dynamical fluctuations than with the

covalent bond. The temporal evolution of the corresponding RMSD is illustrated in Fig. S3 in SI.

Concerning the residue specific analysis, Fig. 4(a) shows IFIE plots of some important residues

whose stabilization energies with the N3 ligand are larger than -10 kcal/mol in the presence of covalent

bond with C145. As can be seen, H163, which is treated as a protonated HIP,[24,34] is a crucial residue with the largest stabilization in both "static" and "dynamic" evaluations. Other notable stabilizing residues are M165, E166, L167 and Q189. The important contributions from these five residues were consistently remarked in Refs. 14, 17-21 and 24 as well. A considerable amount of decrease in stabilization (ca. 15 kcal/mol) due to dynamical motion is found for H163, while those of M165, E166, L167 and Q189 are relatively minor. These relations may be explained[34] by the differences in the contact distances between the residues and the N3 ligand. Kneller et al.[53] investigated the structural changes of 6LU7 by MD simulation and observed that a loop consisting of E166-G170 in M^{pro} moves away from the N3 ligand. The present results illustrating the lower contributions of M165, E166 and L167 are consistent with their observation.

Figure 5(a) illustrates the dynamically averaged PIEDA values of the leading residues in the interactions between M^{pro} and N3 in the presence of covalent bond with C145. The electrostatic (ES) term is the main source of stabilization associated with hydrogen bonds, and its amount for H163 is the largest one among all the residues. In comparison with the "static" PIEDA results in Ref. 24, the contributions from the charge transfer (CT) term are decreased. In exchange for CT, the dispersion (DI) term plays a supplementary role in stabilizing the dynamical ligand binding, especially for M165, E166 and Q189, thus providing useful information for drug design.

As we have seen above, there are several differences between the "static" and "dynamic" structures in interactions between M^{pro} and N3. We analyzed the correlation between "dynamic" IFIE values and interacting distances and found[34] that the correlation coefficients have substantial values of 0.6-0.7 except for some cases. This fact implies that thermal fluctuations under physiological cellular condition certainly affect the interactions between the N3 ligand and surrounding residues.

Figures 4(b) and 5(b) exhibit the results of IFIE and PIEDA in the absence of the covalent bond[24] between the N3 ligand and C145 of M^{pro}. The effects of dynamic reduction in ligand-residue interactions are more significant in the absence than in the presence of the covalent bond, especially for H163, E166 and M165. The strong ES attraction between H163 and N3 is greatly reduced due to the large fluctuation and the position shift of the ligand in the pharmacophore, as seen in Fig. 3(b). The IFIE values of E166 also differ between the static and dynamic configurations substantially due to the flexibility of ligand position. The PIEDA results in Fig. 5(b) indicate that the important ES contributions from L167 and Q189, and DI contributions from M165, Q189 and H41 remain even in the presence of dynamic fluctuations of N3 ligand.

Principal component analysis (PCA).

We first show the results for the PCA in the presence of the covalent bond to C145. The important residues that were selected according to the PCA weight C_j defined by eq. 4 are Q189, E166, A191,

H163 and L167, which are consistent with those enumerated in the previous work[34] (see also Fig.

4(a)). It is here noted that the result of the selection is the same even when the summation over the

PCA mode i is confined to $i \leqq 10$ in eq. 4. Figure 6 illustrates the eigenvalue λ_i of each PCA

component, thus showing that the contributions from the top 10 components (PC1 – PC10) cover about

80 % of the total sum. Figure 7 then depicts the important residues with high magnitudes (absolute

values) in each eigenvector of PC1 – PC5. Q189 shows a dominant contribution in PC1, and E166 in

PC2; in these two mode components, each single residue (Q189 and E166) alone retains an essential

interaction with the ligand N3. In contrast, in the cases of PC3 – PC5, multiple residues collectively

contribute to form the eigenvectors; for example, A191, Q192, L167 and H163 cooperatively interact

with the ligand in the PC3 mode. We thus find an instance that the ligand is trapped via many site

interactions by multiple residues.

As for the case in the absence of the covalent bond to C145, on the other hand, we can enumerate

the essential residues that strongly interact with the ligand N3 based on eq. 4 as Q189, M165, H163,

A191 and H41, which is again consistent with the previous result.[34] As observed in Figure 8, the

contributions from the eigenvalues λ_i of the top 6 components (PC1 – PC6) exceed 80 % of the total

sum. Figure 9 then depicts the important residues with high magnitudes in each eigenvector of PC1 –

PC5. Q189 and M165 give important contributions in PC1, and M165, Q189, H163 and A191 play

vital roles in PC2. A191, M165, A193, Q192 and S46 are the crucial residues in PC3, and similarly

multiple residues cooperatively work for binding in PC4 and PC5 as well. Compared to the case in the presence of the covalent bond to C145, it is characteristic that multiple residues contribute to these interaction modes of PC1 – PC5 more collectively. This dynamical cooperativity in the ligand-residue interactions can be attributed to the larger fluctuations of ligand position (Fig. 3(b)) due to the absence of the covalent bond to C145, thus demonstrating that the ligand N3 is sustained through many site interactions by multiple residues.

In the absence of the covalent bond between the ligand N3 and C145, it is observed that there is a significant change in the ligand binding conformation at around 30 ns in the MD simulation (see Fig. S3 for RMSD). We have therefore attempted two individual PCAs employing the MD trajectories for 0-30 ns (stage 1) and 30-100 ns (stage 2). For the former stage 1, we have found the important residues as Q189, H163, M165, E166 and H41 based on eq. 4. On the other hand, Q189, A191, S46, H41 and M165 have been detected for the latter stage 2. It is noted that S46 listed in the latter was not remarked in the previous study;[34] hence, S46 is a new contributor in the ligand-residue interactions discovered only after the conformation change. It is also noted regarding the relative importance of the residues between the two stages that the weight has increased for Q189, S46, A191, C145 and E47, and decreased for H163, E166, M165, A193 and S144 in the stage 2. The PCA results for 0-30 ns and 30-100 ns are demonstrated in Fig. S4 and Fig. S5, respectively, thus showing the crucial residues for collective interactions with the N3 ligand before and after the conformation change.

Singular value decomposition (SVD).

As mentioned in MATERIALS AND METHODS section, we can obtain two species of eigenvectors, U and V, in the directions of time and amino-acid residue through SVD. Of special interest here are the temporal eigenvectors, which can describe the temporal characteristics in the ligand-residue interactions. Figure 10 illustrates the first, second and third temporal eigenvectors in the presence of the covalent bond between the ligand and C145, thus showing homogeneous, randomized features of temporal variations. In contrast, in the absence of the covalent bond to C145, we observe in Figure 11 that a significant change takes place at around 30 ns, as remarkably detected in the second and third temporal eigenvectors, thus demonstrating the appearance of dynamical "phase transition" associated with the conformation change of the N3 ligand. This detection of the temporal characteristics shows a prominent ability of the present SVD method (see also below). Figure 12 illustrates the distribution of the eigenvalues both in the presence and absence of the covalent bond to C145.

The important residues in the ligand binding can also be enumerated by SVD in terms of the eigenvectors for the amino-acid residues. According to the SVD weight D_j defined by eq. 6, we find H163, Q189, L167, M165 and E166 to be important residues (the result is the same even when the summation over the SVD mode i is confined to $i \leqq 10$) in the presence of the covalent bond to C145,

which is consistent with the previous work.[34)] Table 1 then shows the top five residues with higher

magnitudes in each eigenvector of SV1-SV5. H163, L167, M165, Q189 and H41 are enumerated in

the first eigenvector (SV1) with the magnitudes higher than 0.3. As for the second eigenvector (SV2),

the contributions of Q189 and E166 are dominant, whose magnitudes are higher than 0.2.

In the absence of the covalent bond between the ligand and C145, we find Q189, L167, M165,

H41 and H163 important, as shown in Table 1. Concerning the top five residues with higher

magnitudes in each eigenvector, we observe Q189, L167, H41, H163 and E166 for SV1, Q189, M165,

H163, L167 and E166 for SV2, and Q189, M165, H41, A191 and L167 for SV3. Those residues such

as Q189, M165, H163, E166 and A191 contained in SV2 and SV3 correspond to the residues whose

PCA weights have significantly changed at around 30 ns (see the PCA section above). We thus

consider them to be the crucial residues for the conformation change. Here, we note that more detailed

inspection of the temporal eigenvectors of SV2 and SV3 in Fig. 11 also shows that a temporally broad

transition is occurring between 20 and 40 ns (see also Fig. S3 in SI), where the start and end of the

transition are detected by SV3 and SV2, respectively, as the temporally global change in the sign of

eigenvector components. Besides, it is found that all the top five residues for the first to fifth

eigenvectors (SV1 – SV5) in Table 1 have the magnitudes higher than 0.2; this fact shows again that

the N3 ligand is sustained through many site interactions by multiple residues in the case of the absence

of the covalent bond to C145. It is also characteristic in this case that the important contributions from

S46 appear in the SV4 and SV5, which has been noted in the time-divided PCA above as well.

Comparison to other investigations.

Kneller et al.[53] performed the room temperature X-ray structure analysis of unliganded SARS-CoV-

2 3CL M^{pro} and compared the result with the low-temperature ligand-free and N3-bound structures

and also with the ligand-free MD structures at room temperature. Through these analyses, they found

the significant structural plasticity of M^{pro} in the vicinity of the active site such as the small helix near

P2 group containing residues 45-50, the β-hairpin loop near P3-P4 region with residues 165-170 and

the P5 loop spanning residues 190-194. The dynamic flexibility of these structural regions can account

for the dynamical IFIE interactions observed in the present PCA and SVD analyses.

Through the detailed QM/MM simulations for the complex of SARS-CoV-2 M^{pro} with N3 and its

analogues, Arafet et al.[54] enumerated the important interactions between the M^{pro} residues and the

covalent or non-covalent inhibitors. In addition to the active interactions associated with C145 and

H41, they found vital (primarily hydrogen bond) interactions between the ligand and those residues

such as F140, N142, G143, H163, H164, E166, Q189, T190 and Q192. The collectivity of the

interaction networks is a significant finding in their research, which corresponds well to that in the

present study.

Through the pharmacophore modeling and $1\,\mu$s MD simulations for SARS-CoV-2 M^{pro} in complex with three drug candidates (2A5I, 2OP9 and Indinavir), Yoshino et al.[21] identified the key interactions between the possible inhibitors and the surrounding residues. They thus found that primarily H41, G143 and E166 (further possibly M49, S144, C145, M165, P168 and Q189) formed crucial interactions with the functional groups common among peptide-like inhibitors, and concluded that these interactions should be important targets for designing potential drugs against SARS-CoV-2 M^{pro} since the inhibitors would dynamically interact with M^{pro} at multiple sites in the pharmacophore. Similar descriptions on the ligand-residue interactions were also given by Nutho et al.[20].

Overall, these structural and simulation studies have provided the pictures for designing possible inhibitors against M^{pro} on the basis of ligand-residue interactions consistent with the present dynamical FMO-IFIE study. The advantage of the present PCA/SVD based methodology is that it can give a comprehensive approach to systematically extracting crucial interactions in terms of *ab initio* electron-correlated FMO method for the whole protein-ligand systems with the aid of highly parallelized capacity computing. Through inclusion of explicit water molecules, the present FMO approach could also account for the effects associated with the hydrogen bond networks[20,39] formed by surrounding hydration environments, whose detailed analysis would be an interesting research problem in forthcoming studies.

CONCLUSIONS

In the present work we have performed a combined simulation based on classical MD and *ab initio* FMO methods for the analysis of ligand binding in the SARS-CoV-2 M^{pro}-N3 complex system, thus finding the importance of statistical evaluation of interactions by employing a thousand of dynamically fluctuating structures. Recent studies[17,20-23,53,55-56] concerning SARS-CoV-2 proteins have shed light on the importance of "dynamic" aspects, and our quantum-chemical work accords with this direction. As demonstrated here, the combination of MD and FMO approaches employing a thousand of sampled structures has now been made feasible by using massive computing resources such as the Fugaku supercomputer. The present approach can thus provide a computational way to satisfy the demands that the ligand-residue interactions should be described with quantum chemical accuracy and in a dynamical manner, while the accomplishment of fully *ab initio* molecular dynamics would be an ultimate goal.

As novel computational tools in statistical analysis, we have proposed in this study the use of PCA and SVD to systematically elicit the dynamical characteristics in ligand-residue IFIEs. Applied to the SARS-CoV-2 M^{pro}-N3 system, we have found dynamically correlated, collective interactions between ligand and surrounding residues, whose behaviors are mostly consistent with those observed in the literature but containing some new findings. We consider this proposed methodology will extend the conventional FMO-based "static" analyses of ligand binding such as VISCANA[57] and similar

approaches[37,58] to the "dynamic" ones, thus enabling the dynamical descriptions of interaction fingerprint to represent the interactions between ligand and multiple residues for pharmaceutical applications. Furthermore, we can anticipate more general utility of the present method such as the application to the dynamical descriptions of networks composed of hydrogen bonding[39] and dispersion interactions in molecular systems, which are essential for molecular designs including drug discovery. For instance, we may think of the design of inhibitors that can utilize the dynamical interactions with multiple sites of target protein.

Supporting Information

Fragmentation scheme of N3 ligand, RMSDs for MD simulations in the presence and absence of the covalent bond between N3 and C145, and PCA results for stages 1 and 2 in the absence of the covalent bond

Corresponding Author

*E-mail: tanaka2@kobe-u.ac.jp

ORCID

Shigenori Tanaka: 0000-0002-6659-2788

Koji Okuwaki: 0000-0002-4510-5717

Kaori Fukuzawa: 0000-0001-5357-8250

Yuto Komeiji: 0000-0002-9708-6175

Yoshio Okiyama: 0000-0002-6775-8604

Yuji Mochizuki: 0000-0002-7310-5183

Notes

ACKNOWLEDGMENTS

REFERENCES

1) K.M. Merz, Jr., D. Ringe, C.H. Reynolds, ed., "Drug Design: Structure- and Ligand-based Approaches", Cambridge University Press, 2010.

2) S. Forli, R. Huey, M.E. Pique, M.F. Sanner, D.S. Goodsell, A.J. Olson, Nature Prot. 11, 905 (2016).

3) F.D. Prieto-Martinez, M. Arciniega, J.L. Medina-Franco, TIP Rev Espec en Ciencias Quemico-Biologicas 21, 65−87 (2018).

4) S. Uehara and S. Tanaka, Molecules 21, 1604 (2016).

5) L.A. Defelipe, J.P. Arcon, C.P. Modenutti, M.A. Marti, A.G. Turjanski, X. Barril, Molecules 23, 3269 (2018).

6) M. Araki, N. Kamiya, M. Sato, M. Nakatsui, T. Hirokawa, Y. Okuno, J. Chem. Inf. Model. 56, 2445-2456 (2016).

7) S. Uehara and S. Tanaka, J. Chem. Inf. Model. 57, 742-756 (2017).

8) Z. Liu, M. Su, L. Han, J. Liu, Q. Yang, Y. Li, R. Wang, Acc. Chem. Res. 50, 302−309 (2017).

9) D.L. Mobley, P.V. Klimovich, J. Chem. Phys. 137, 230901 (2012).

10) Z. Deng, C. Chuaqui, J. Singh, J. Med. Chem. 47, 337-344 (2004).

11) V.P. Waman, N. Sen, M. Varadi, A. Daina, S.J. Wodak, V. Zoete, S. Velankar, C. Orengo, Briefings in Bioinformatics, 00(00), 1–27 (2020). doi: 10.1093/bib/bbaa362

12) https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (access confirmed on 2 February 2021).

13) Z. Xu, C. Peng, Y. Shi, Z. Zhu, K. Mu, X. Wang and W. Zhu, bioRxiv 2020.01.27.921627 (2020).

14) Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu,

X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao and H. Yang, Nature 582, 289 (2020).

15) Z. Li, X. Li, Y.-Y. Huang, Y. Wu, R. Liu, L. Zhou, Y. Lin, D. Wu, L. Zhang, H. Liu, X. Xu, K. Yu, Y. Zhang, J. Cui, C.-G. Zhan, X. Wang and H.-B. Luo, Proc. Natl. Acad. Sci. 117, 27381-27387 (2020).

16) S. K. Enmozhi, K. Raja, I. Sebastine and J. Joseph, J. Biomol. Struct. Dyn. 1 (2020). DOI: 10.1080/07391102.2020.1760136

17) L. Mittal, A. Kumari, M. Srivastava, M. Singh and S. Asthana, J. Biomol. Struct. Dyn. 1 (2020). DOI: 10.1080/07391102.2020.1768151

18) R. Yu, L. Chen, R. Lan, R. Shen and P. Li, Int. J. Antimicrob. Agents 56, 106012 (2020).

19) A. I. Owis, M. S. El-Hawary, D. El Amir, O. M. Aly, U. R. Abdelmohsen and M. S. Kamel, RSC Adv. 10, 19570 (2020).

20) B. Nutho, P. Mahalapbutr, K. Hengphasatporn, N. C. Pattaranggoon, N. Simanon, Y. Shigeta, S. Hannongbua and T. Rungrotmongkol, Biochemistry 59, 1769 (2020).

21) R. Yoshino, N. Yasuo and M. Sekijima, Sci. Rep. 10, 12493 (2020).

22) A. Gahlawat, N. Kumar, R. Kumar, H. Sandhu, I. P. Singh, S. Singh, A. Sjöstedt and P. Garg, J. Chem. Inf. Model. 60, 5781–5793 (2020).

23) D. Suárez and N. Díaz, J. Chem. Inf. Model. 60, 5815–5831 (2020).

24) R. Hatada, K. Okuwaki, Y. Mochizuki, Y. Handa, K. Fukuzawa, Y. Komeiji, Y. Okiyama and S. Tanaka, J. Chem. Inf. Model. 60, 3593 (2020).

25) D.G. Fedorov and K. Kitaura, ed., "The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems", CRC Press, London, 2009.

26) K. Kitaura, E. Ikeo, T. Asada, T. Nakano and M. Uebayasi, Chem. Phys. Lett. 313, 701 (1999).

27) D. G. Fedorov, T. Nagata and K. Kitaura, Phys. Chem. Chem. Phys. 14, 7562 (2012).

28) S. Tanaka, Y. Mochizuki, Y. Komeiji, Y. Okiyama and K. Fukuzawa, Phys. Chem. Chem. Phys. 16, 10310 (2014).

29) Y. Mochizuki, S. Tanaka, and K. Fukuzawa, ed., "Recent Advances of the Fragment Molecular Orbital Method: Enhanced Performance and Applicability", Springer, Singapore, 2021.

30) D. G. Fedorov and K. Kitaura, J. Comput. Chem. 28, 222 (2007).

31) T. Tsukamoto, K. Kato, A. Kato, T. Nakano, Y. Mochizuki and K. Fukuzawa, J. Comput. Chem. Japan 14, 1 (2015).

32) Y. Okiyama, T. Nakano, C. Watanabe, K. Fukuzawa, Y. Mochizuki and S. Tanaka, J. Phys. Chem. B 122, 4457 (2018).

33) Y. Okiyama, C. Watanabe, K. Fukuzawa, Y. Mochizuki, T. Nakano and S. Tanaka, J. Phys. Chem. B 123, 957 (2019).

34) R. Hatada, K. Okuwaki, K. Akisawa, Y. Mochizuki, Y. Handa, K. Fukuzawa, Y. Komeiji, Y. Okiyama, S. Tanaka, Appl. Phys. Express 14, 027003 (2021).

35) https://www.fujitsu.com/global/about/innovation/fugaku/

36) V. Klema, A. Laud, IEEE Trans. Automat. Contr. 25, 164-176 (1980).

37) K. Maruyama, Y. Sheng, H. Watanabe, K. Fukuzawa, S. Tanaka, Comput. Theor. Chem. 1132, 23-34 (2018).

38) S. Tanaka, C. Watanabe, T. Honma, K. Fukuzawa, K. Ohishi, T. Maruyama, J. Mol. Graph. Model. 100, 107650 (2020).

39) S. Matsunaga, Y. Hano, Y. Saito, K.J. Fujimoto, T. Kumasaka, S. Matsumoto, T. Kataoka, F. Shima, S. Tanaka, J. Mol. Graph. Model. 77, 51-63 (2017).

40) Molecular Operating Environment (MOE); Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7 (2013).

41) D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, J. Comput. Chem. 26, 1668 (2005).

42) D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman AMBER 2016, University of California, San Francisco (2016).

43) Y. Komeiji, T. Ishida, D. G. Fedorov and K. Kitaura, J. Comput. Chem. 28, 1750 (2007).

44) K. Fukuzawa, I. Kurisaki, C. Watanabe, Y. Okiyama, Y. Mochizuki, S. Tanaka and Y. Komeiji, Comput. Theor. Chem. 1054, 29 (2015).

45) Y. Komeij, Y. Okiyama, Y. Mochizuki and K. Fukuzawa, Chem-Bio Informatics J. 17, 72 (2017).

46) A. Szabo and N. S. Ostlund, "Modern Quantum Chemistry", MacMillan, New York, 1982.

47) Y. Mochizuki, S. Koikegami, T. Nakano, S. Amari and K. Kitaura, Chem. Phys. Lett. 396, 473 (2004).

48) Y. Mochizuki, T. Nakano, S. Koikegami, S. Tanimori, Y. Abe, U. Nagashima and K. Kitaura, Theor. Chem. Acc. 112, 442 (2004).

49) Y. Mochizuki, K. Yamashita, T. Murase, T. Nakano, K. Fukuzawa, K. Takematsu, H. Watanabe and S. Tanaka, Chem. Phys. Lett. 457, 396 (2008).

50) C. E. Dykstra and E. R. Davidson, Int. J. Quantum Chem. 78, 226 (2000).

51) J. B. Foresman and A. Frisch, "Exploring chemistry with electronic structure methods", 2nd edn., Gaussian Inc., Pittsburgh, 1996.

52) T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi and K. Kitaura, Chem. Phys. Lett. 318, 614 (2000).

53) D. W. Kneller, G. Phillips, H. M. O'Neill, R. Jedrzejczak, L. Stols, P. Langan, A. Joachimiak, L. Coates and A. Kovalevsky, Nat. Commun. 11, 3202 (2020).

54) K. Arafet, N. Serrano-Aparicio, A. Lodola, A. Mulholland, F.V. González, K. Swiderek, V. Moliner, Chem. Sci. (2021) DOI: 10.1039/d0sc06195f

55) M. Macchiagodena, M. Pagliai, M. Karrenbrock, G. Guarnieri, F. Iannone and P. Procacci, J. Chem. Theory Comput. 16, 7160–7172 (2020).

56) P. R. Arantes, A. Saha and G. Palermo, ACS Cent. Sci. 6, 1654 (2020).

57) S. Amari, M. Aizawa, J. Zhang, K. Fukuzawa, Y. Mochizuki, Y. Iwasawa, K. Nakata, H. Chuman and T. Nakano, J. Chem. Inf. Model. 46, 221 (2006).

58) R. Kurauchi, C. Watanabe, K. Fukuzawa, S. Tanaka, Comput. Theor. Chem. 1061, 12-22 (2015).

59) https://www.r-ccs.riken.jp/en/topics/fugaku-coronavirus.html

Table

Table 1. Top five residues with higher magnitudes (denoted in parentheses) in each eigenvector for SV1-SV5 in SVD analysis, along with those by the total weight, eq. 6. The results for the presence and absence of the covalent bond with C145 are shown in upper and lower panels, respectively.

Total	H163	Q189	L167	M165	E166

SV1	H163 (0.489)	L167 (0.412)	M165 (0.397)	Q189 (0.384)	H41 (0.319)
SV2	Q189 (0.815)	E166 (-0.499)	H163 (-0.169)	L167 (-0.144)	T190 (-0.094)
SV3	E166 (0.764)	Q189 (0.399)	M165 (-0.314)	H163 (-0.236)	L167 (-0.176)
SV4	A191 (-0.759)	H163 (0.420)	Q192 (0.380)	L167 (-0.225)	P168 (0.108)
SV5	L167 (0.558)	A191 (-0.482)	A193 (0.383)	Q192 (-0.347)	H163 (-0.206)

Total	Q189	L167	M165	H41	H163

SV1	Q189 (0.572)	L167 (0.510)	H41 (0.335)	H163 (0.269)	E166 (0.259)
SV2	Q189 (0.641)	M165 (-0.541)	H163 (-0.368)	L167 (-0.233)	E166 (-0.204)
SV3	Q189 (0.495)	M165 (0.462)	H41 (-0.390)	A191 (-0.363)	L167 (-0.287)
SV4	A191 (-0.717)	M165 (-0.386)	S46 (0.269)	H41 (0.257)	A193 (0.241)
SV5	H163 (-0.584)	Q192 (0.362)	E166 (0.340)	S46 (0.288)	N142 (0.219)
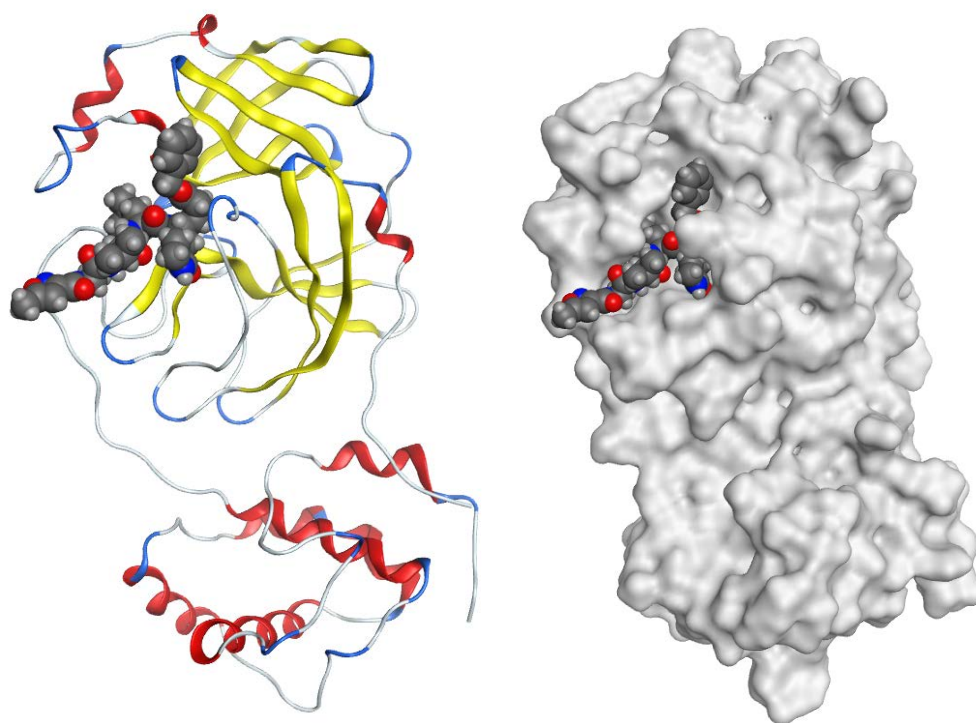
Figure 1. Structure of M^pro-N3 complex (PDB entry: 6LU7) by two ways in which M^pro is depicted by ribbon (left) and molecular surface (right) representations. The N3 ligand is depicted by the sphere representation.
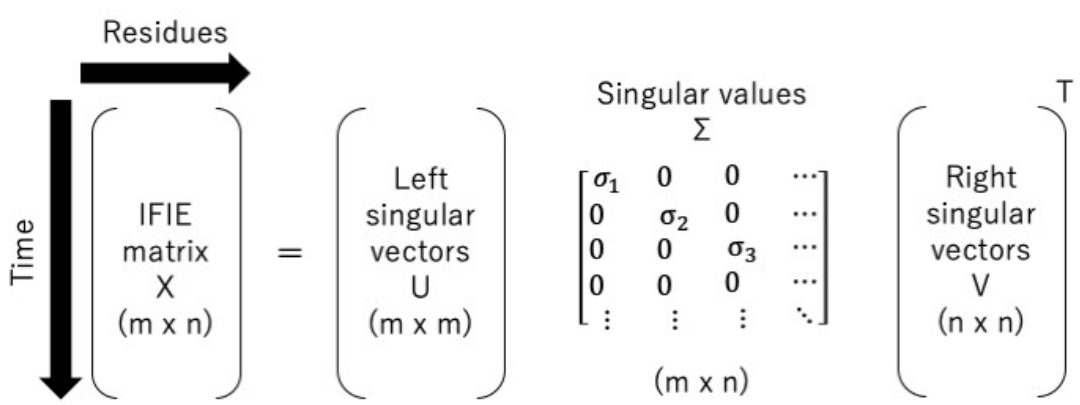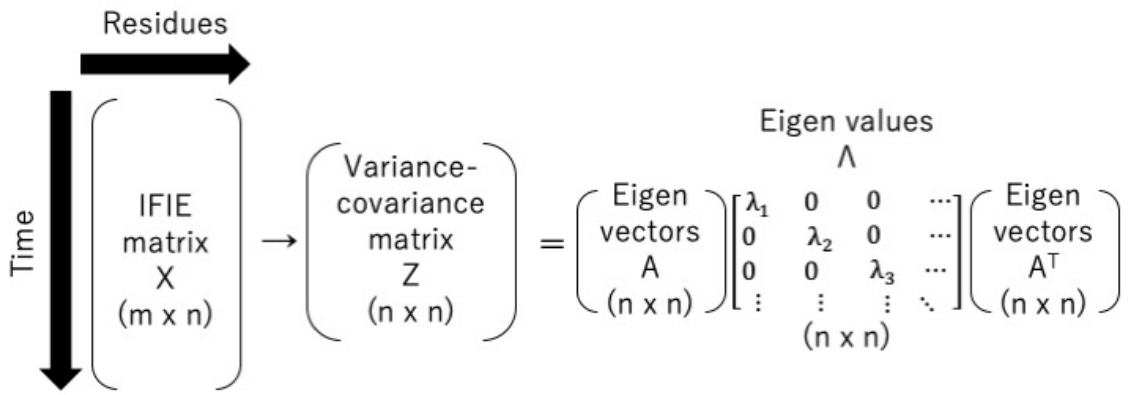
Figure 2. Schematic representations for PCA (upper panel) and SVD (lower panel).
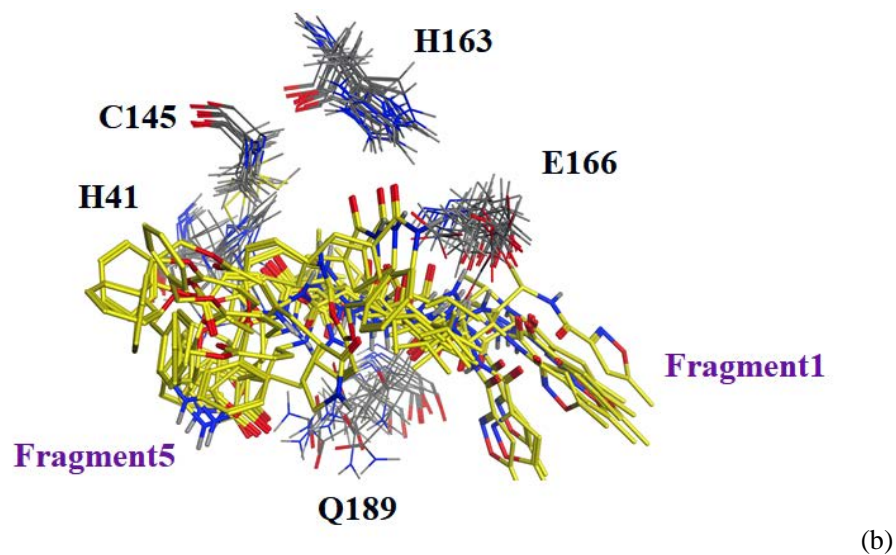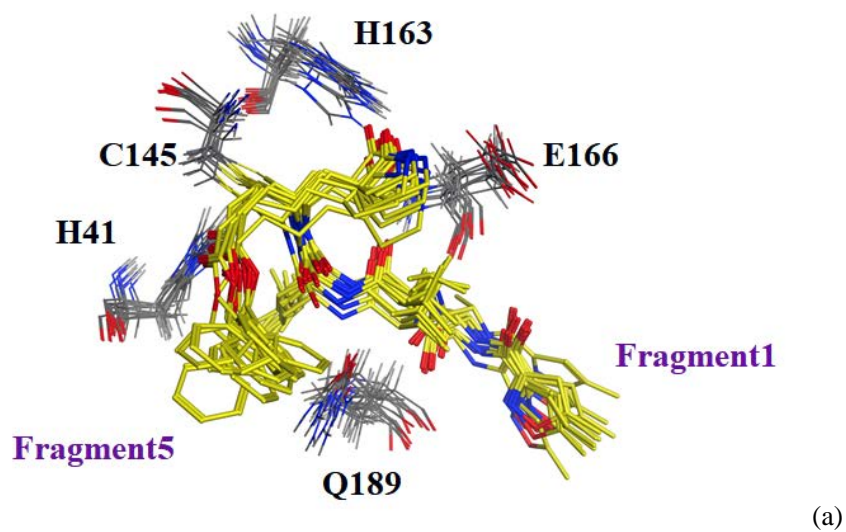
Figure 3. Superimposed view of snapshots from the 100 ns MD trajectory sampled with 10 ns intervals for the N3 ligand and crucial residues in pharmacophore. (a) The N3 ligand which is covalently bonded to C145 is depicted in yellow color. (b) The N3 ligand depicted in yellow is not covalently bonded to C145.
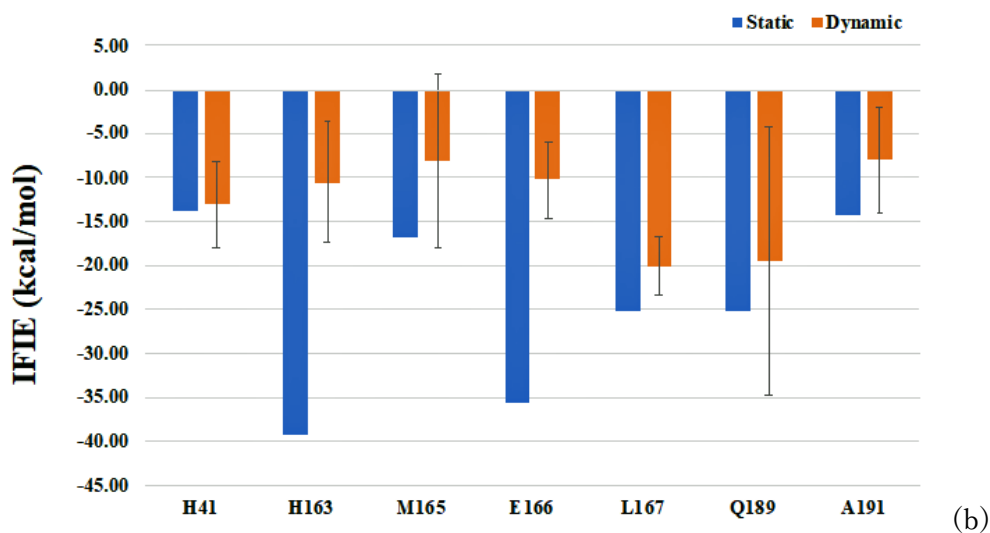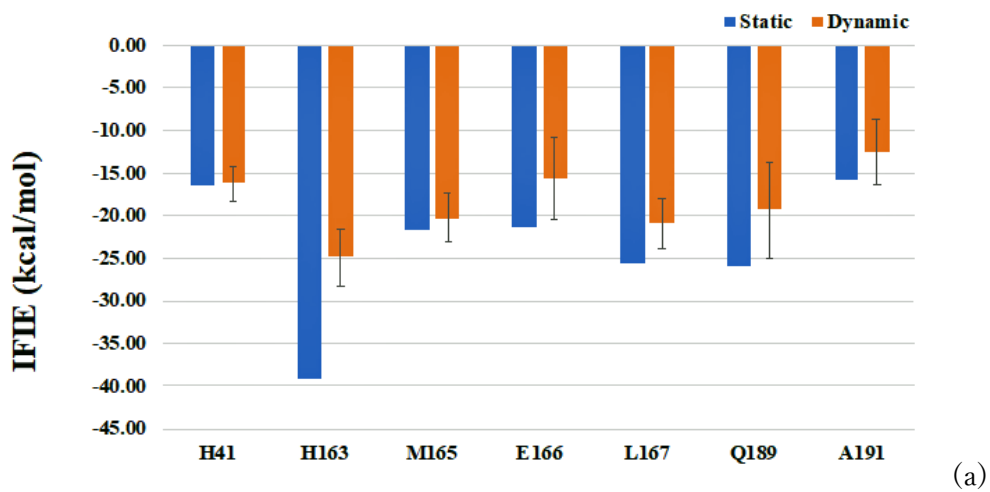
Figure 4. IFIE values of leading residues interacting (larger than -10 kcal/mol of total stabilization) with M^{pro} in the cases of the presence (a) and absence (b) of the covalent bond between N3 and C145. Blue and orange bars correspond to the single value by "static" structure and the averaged value by a thousand of "dynamic" structures, respectively. Vertical line for the latter refers to the standard deviation.
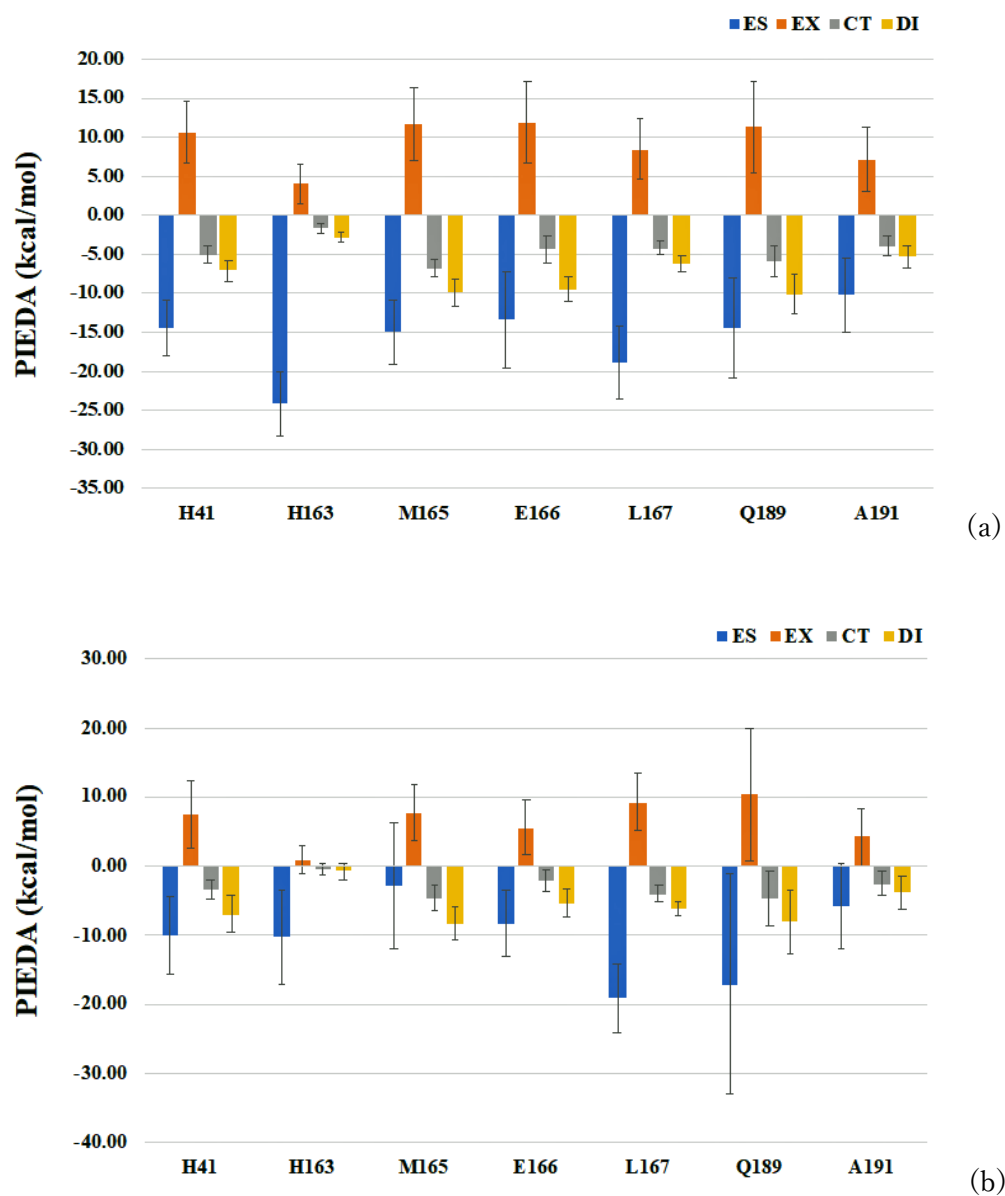
Figure 5. Averaged PIEDA values of crucial residues interacting (more than -10 kcal/mol of total stabilization) with M^{pro} over a thousand of "dynamic" structures in the cases of the presence (a) and absence (b) of the covalent bond between N3 and C145. Blue, orange, gray and yellow bars correspond to ES, EX, CT and DI terms, respectively. Vertical line refers to the standard deviation.
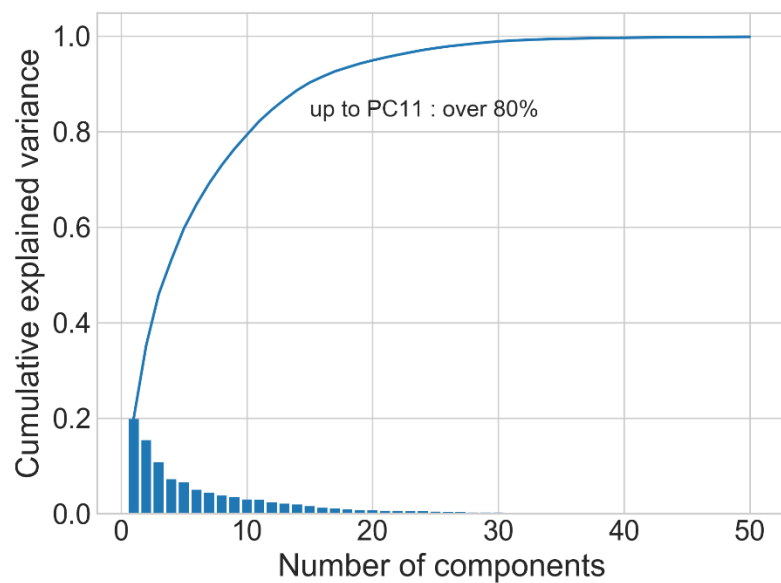
Figure 6. Distribution of eigenvalues in PCA in the presence of the covalent bond to C145. Solid curve represents the cumulative contribution.
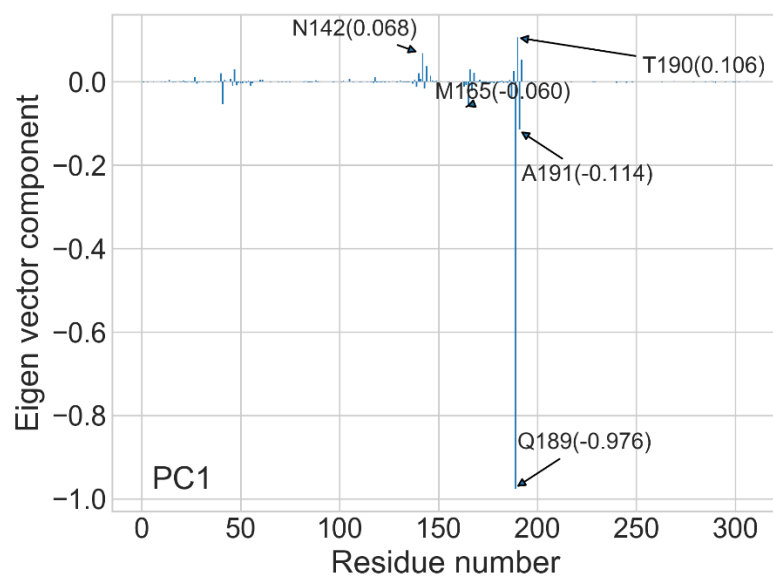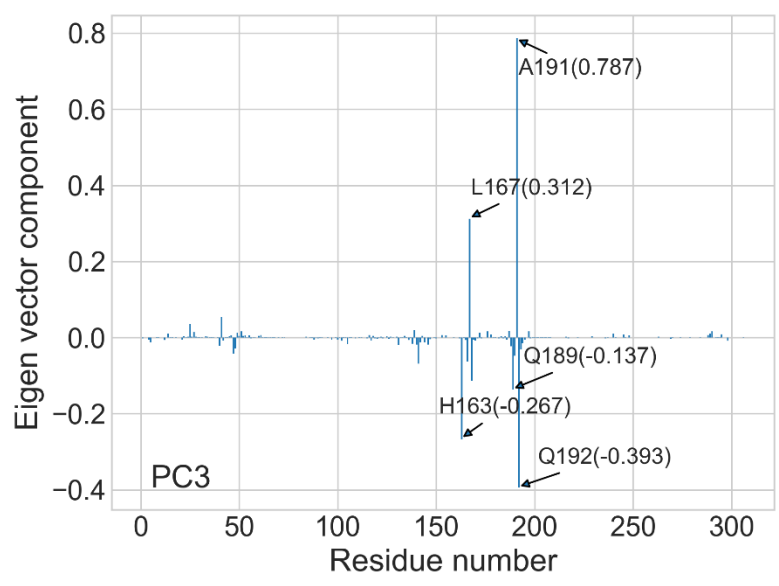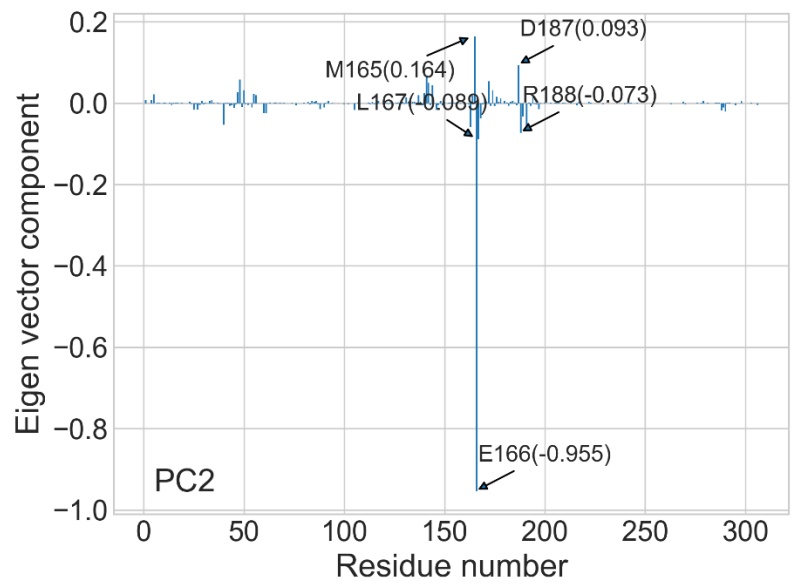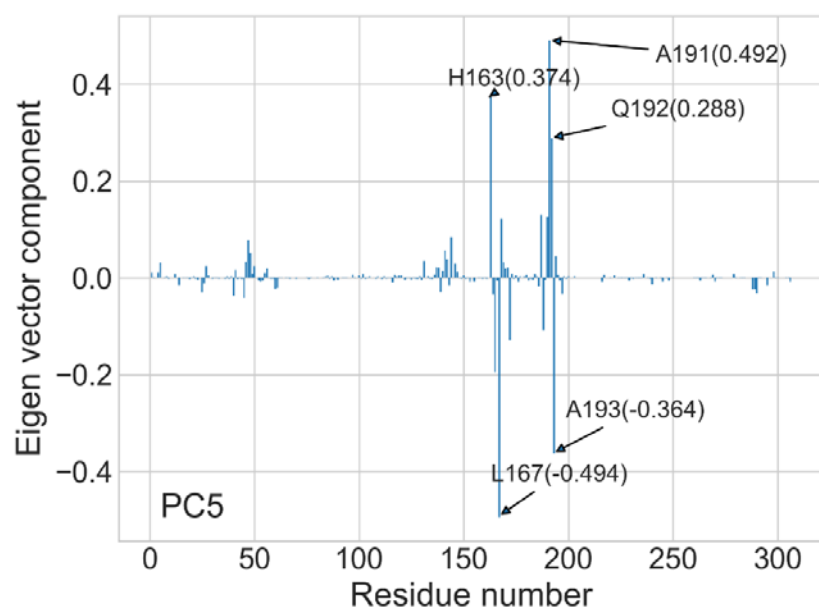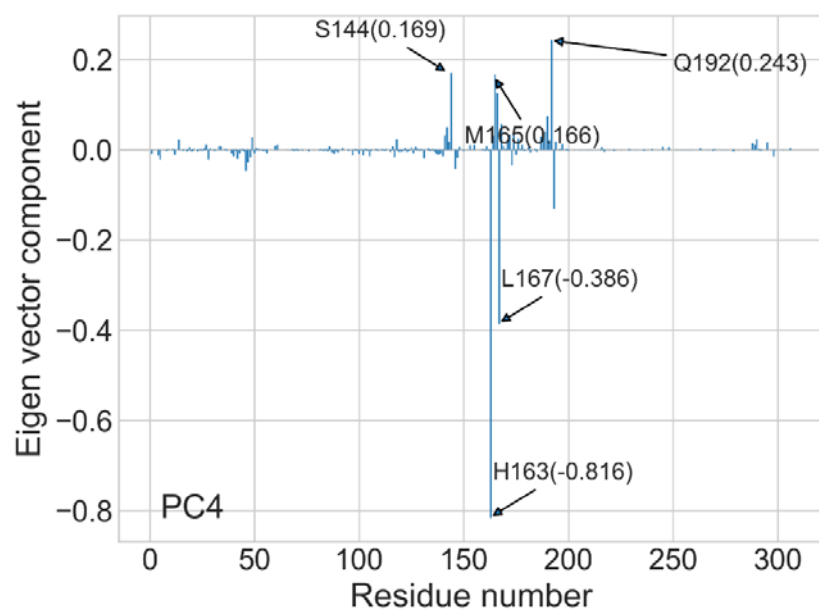
Figure 7. Residue components for the eigenvectors of PC1-PC5 in PCA in the presence of the covalent

bond with C145.
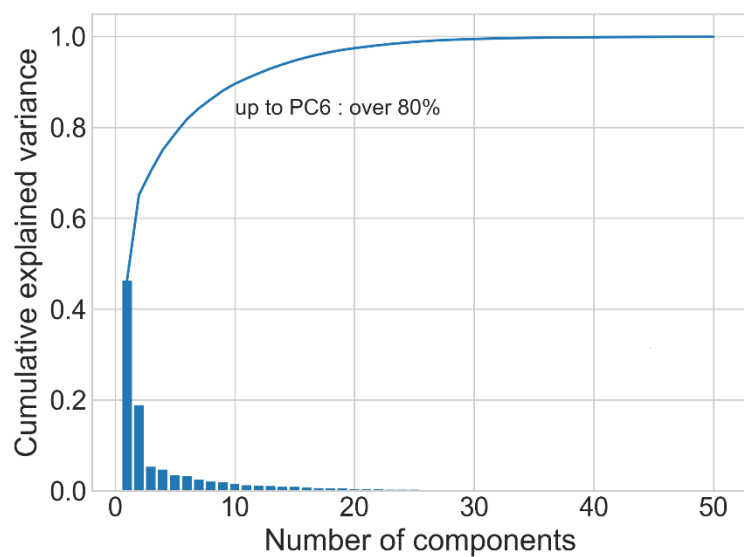
Figure 8. Distribution of eigenvalues in PCA in the absence of the covalent bond to C145. Solid curve represents the cumulative contribution.
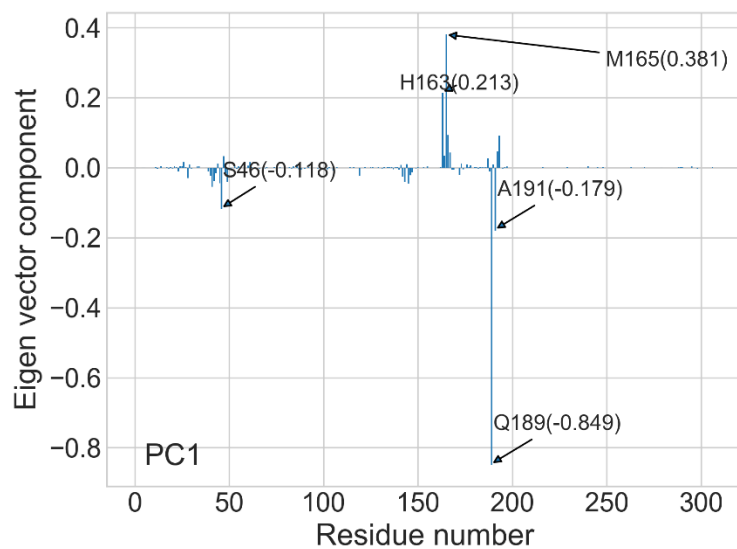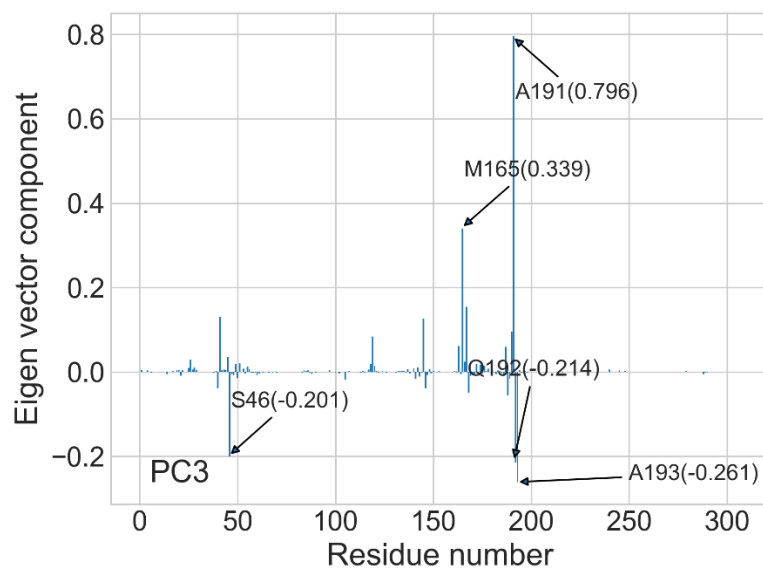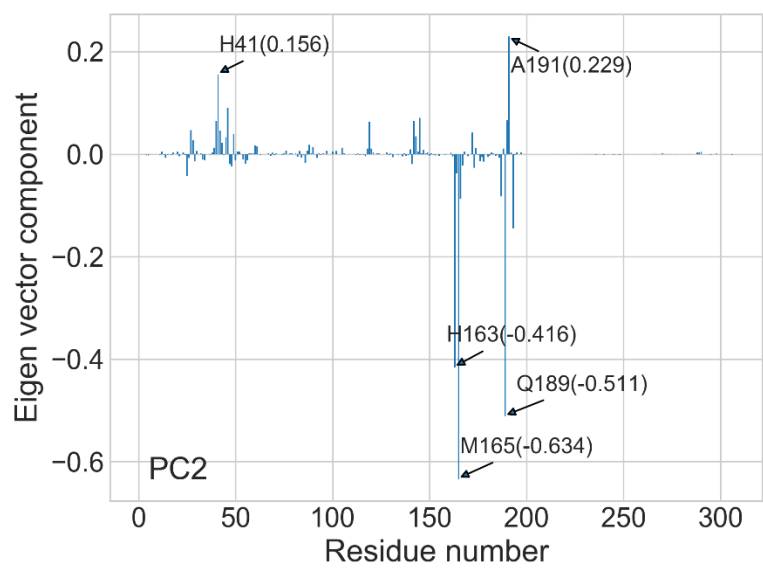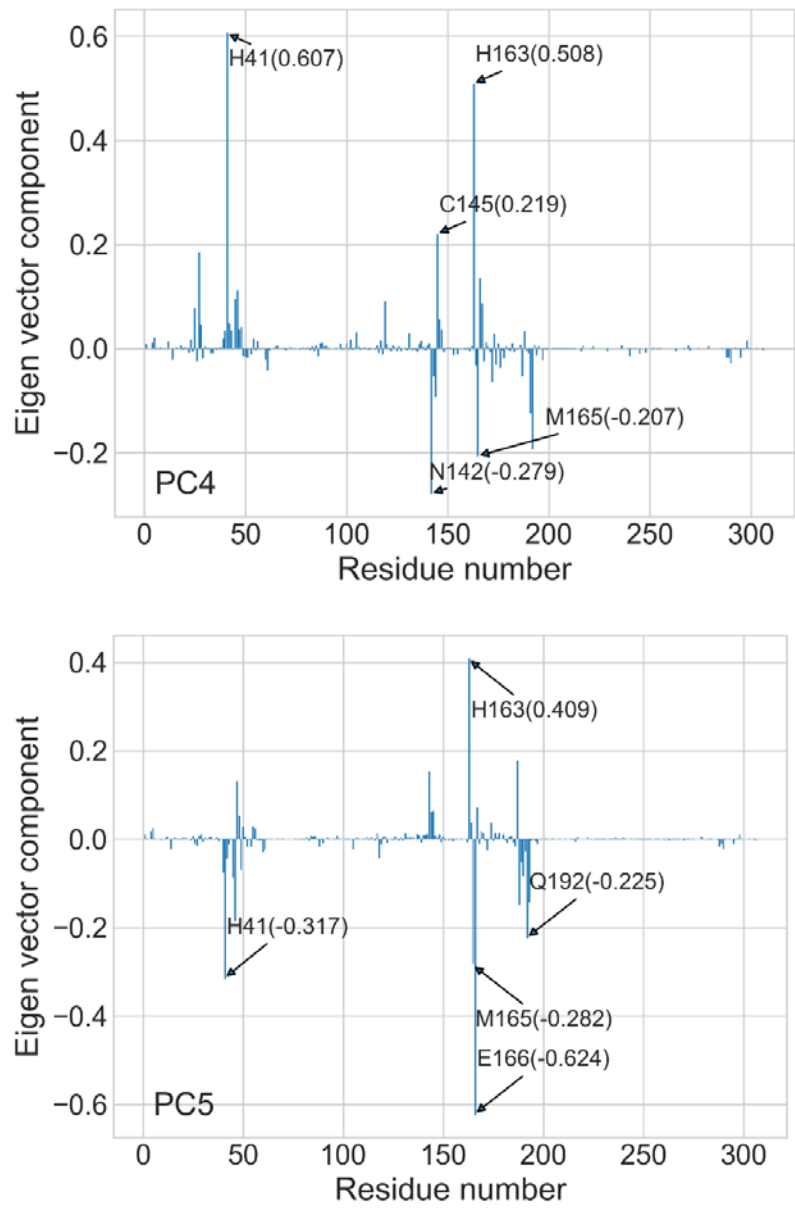
Figure 9. Residue components for the eigenvectors of PC1-PC5 in PCA in the absence of the covalent bond with C145.
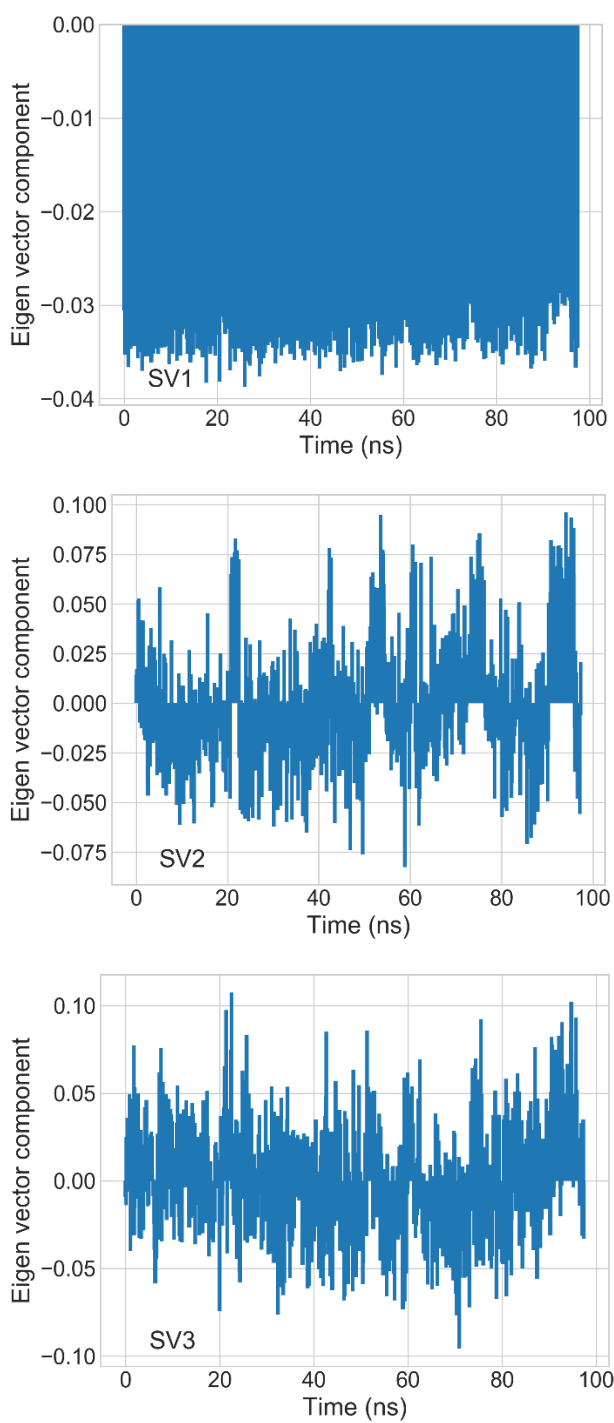
Figure 10. Temporal components (U) for the eigenvectors of SV1-SV3 in SVD in the presence of the covalent bond to C145.
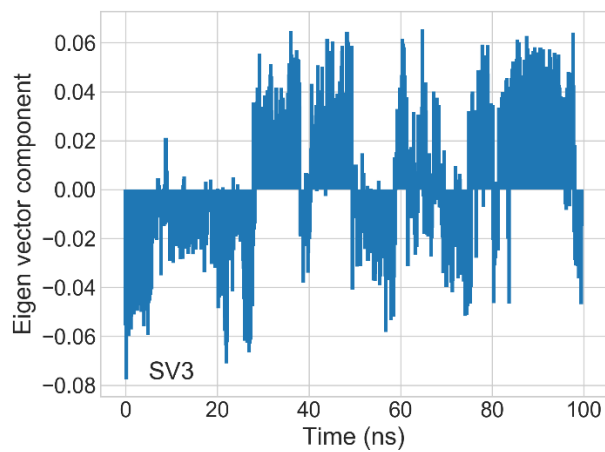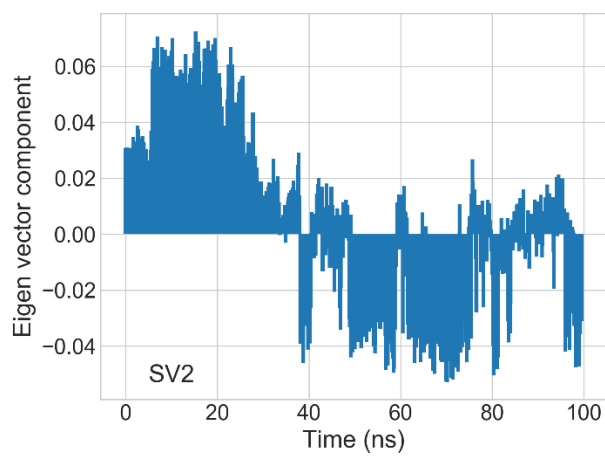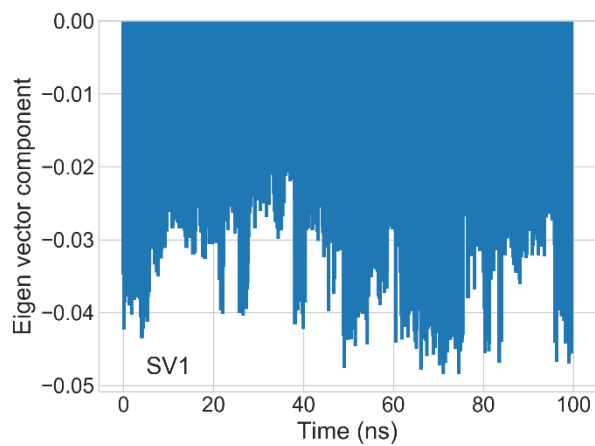
Figure 11. Temporal components (U) for the eigenvectors of SV1-SV3 in SVD in the absence of the covalent bond to C145.

(a)

(b)
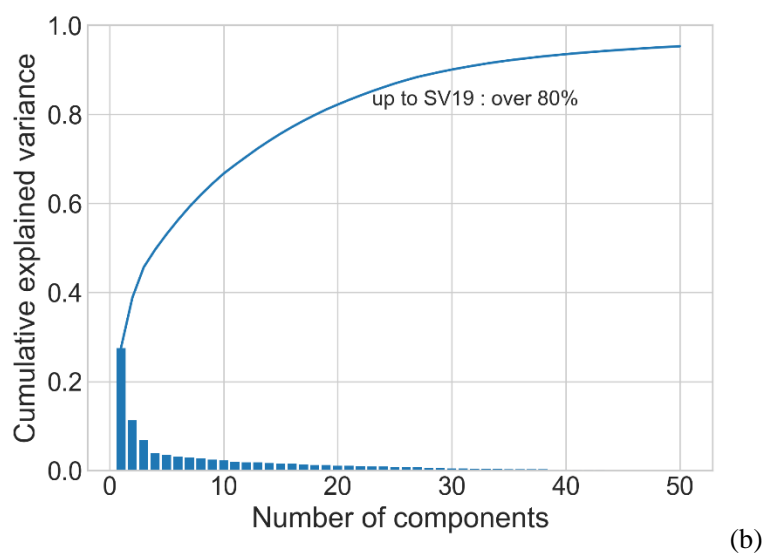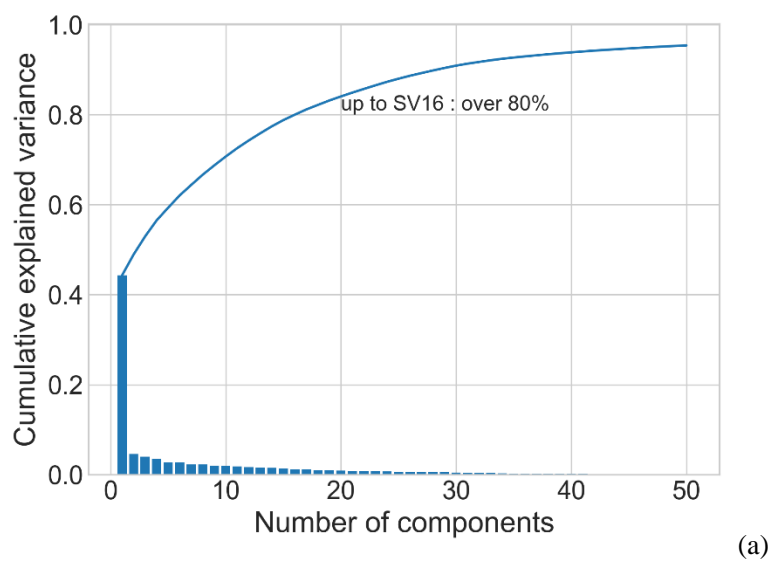
Figure 12. Distributions of eigenvalues in SVD in the presence (a) and absence (b) of the covalent bond to C145. Solid curves represent the cumulative contributions.

(Supporting Information)

Dynamical Cooperativity of Ligand-Residue Interactions Evaluated with the Fragment Molecular Orbital Method

Shigenori Tanaka[1*], Shusuke Tokutomi[1], Ryo Hatada[2], Koji Okuwaki[2], Kazuki Akisawa[2], Kaori Fukuzawa[3,4,5], Yuto Komeiji[6], Yoshio Okiyama[7], Yuji Mochizuki[2,5]

[1] *Graduate School of System Informatics, Department of Computational Science, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe 657-8501, Japan*

[2] *Department of Chemistry and Research Center for Smart Molecules, Faculty of Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan*

[3] *School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa-Ku, Tokyo 142-8501, Japan*

[4] *Department of Biomolecular Engineering, Graduate School of Engineering, Tohoku University, 6-6-11 Aoba, Aramaki, Aoba-ku, Sendai 980-8579, Japan*

[5] *Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan*

[6] *Biomedical Research Institute, AIST, Tsukuba Central 6, Tsukuba, Ibaraki 305-8566, Japan*

[7] *Division of Medicinal Safety Science, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa 201-9501, Japan*

In this Supporting Information, a number of supplementary figures addressed in the main text are shown with their legends.
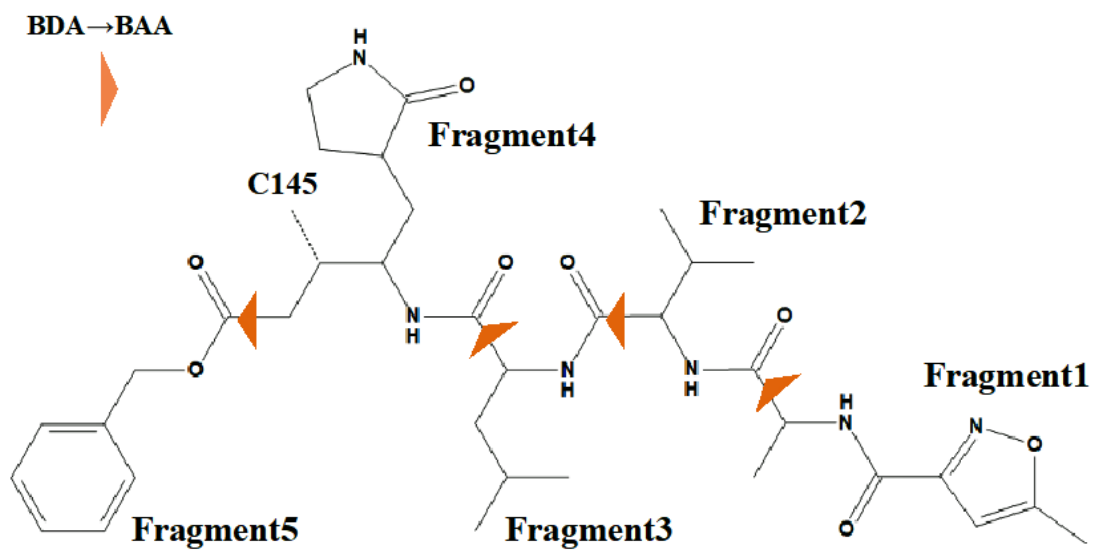
Figure S1. Fragmentation of N3 ligand. The fragmentation point is identified by orange triangle indicating the direction of bond detachment atom (BDA) to bond attachment atom (BAA). There is a covalent bond to C145.
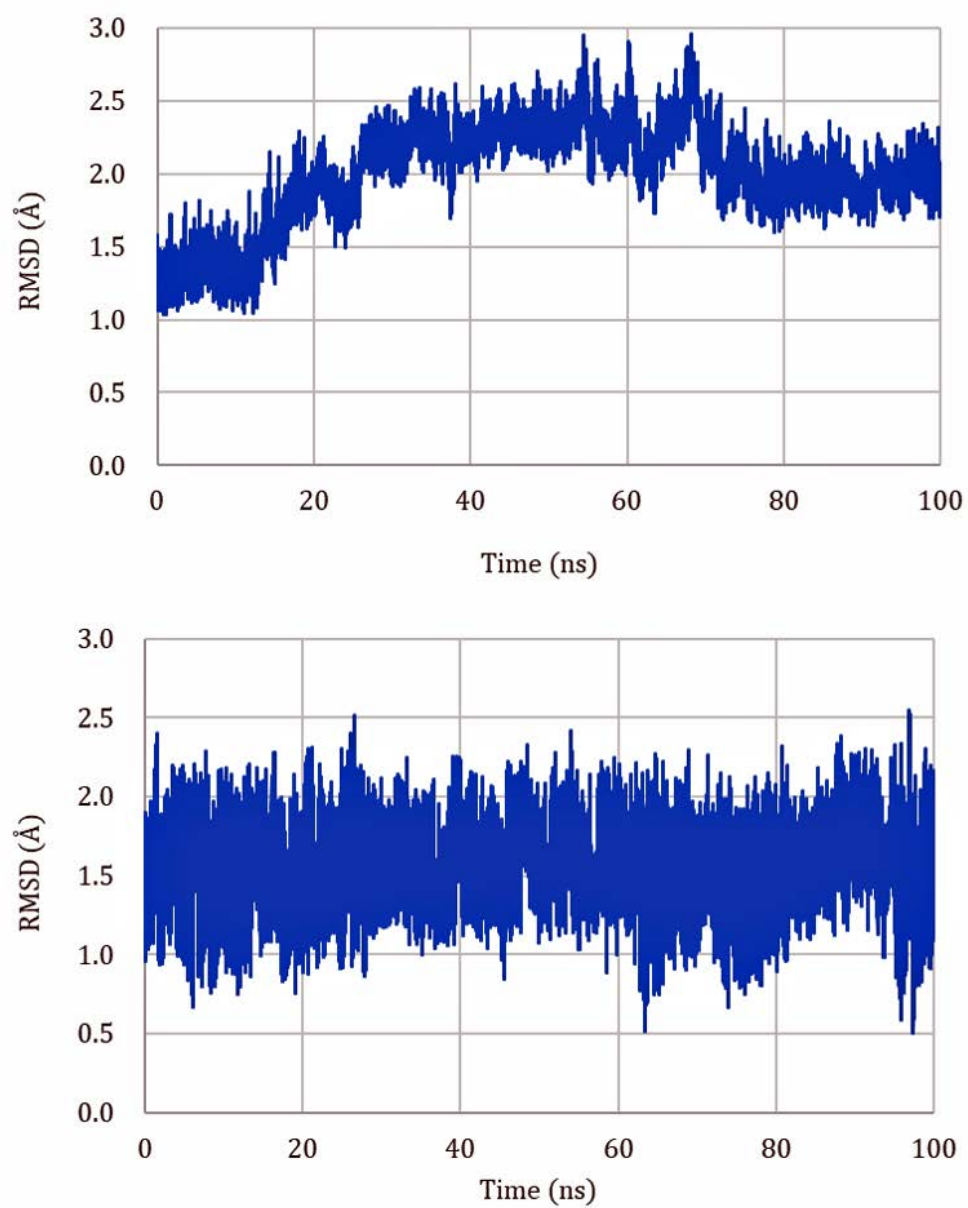
Figure S2. RMSDs for 100 ns MD in the presence of the covalent bond between N3 ligand and C145 residue of M^{pro}. Upper panel: C$_\alpha$ atoms in M^{pro}. Lower panel: C and N atoms in N3.
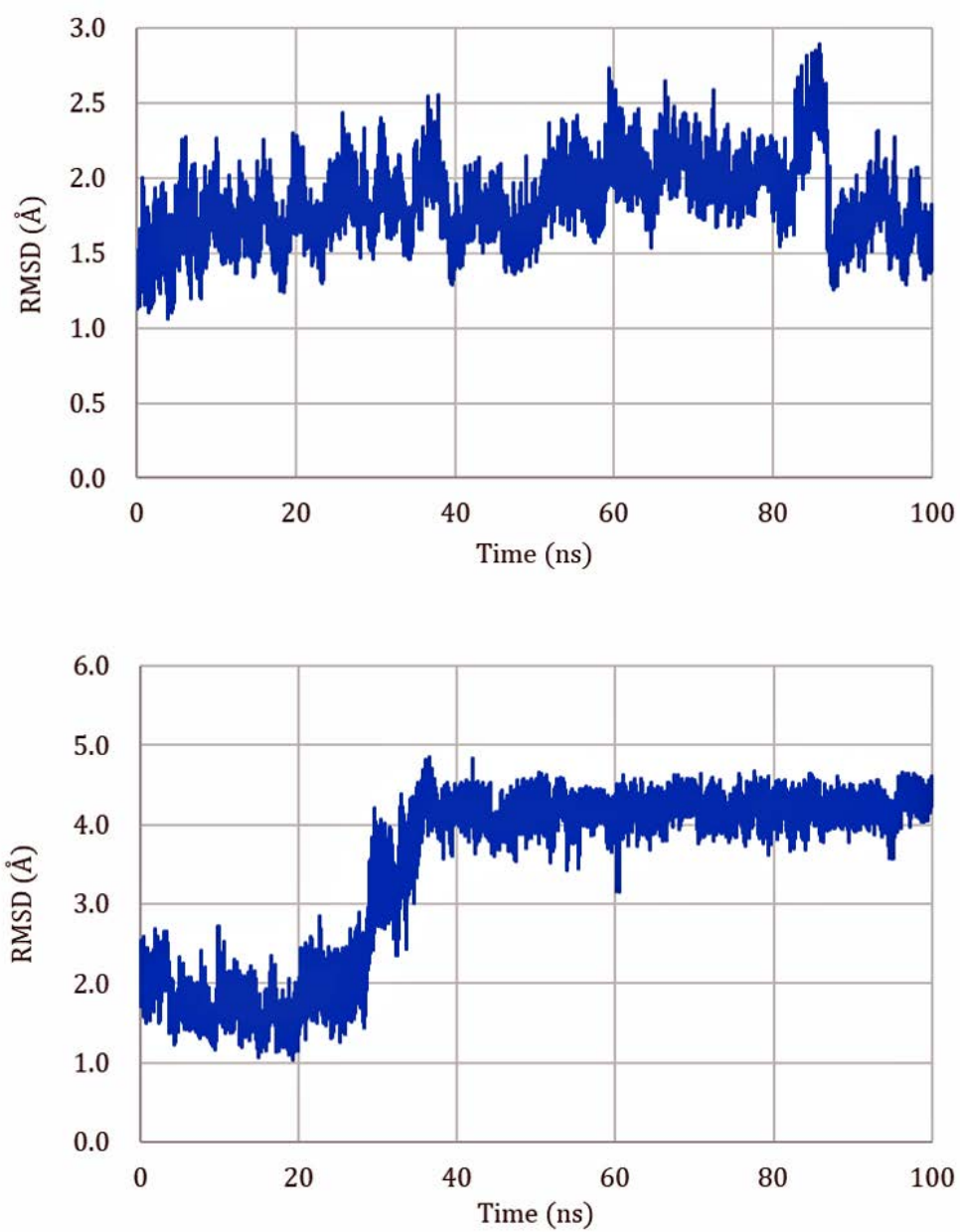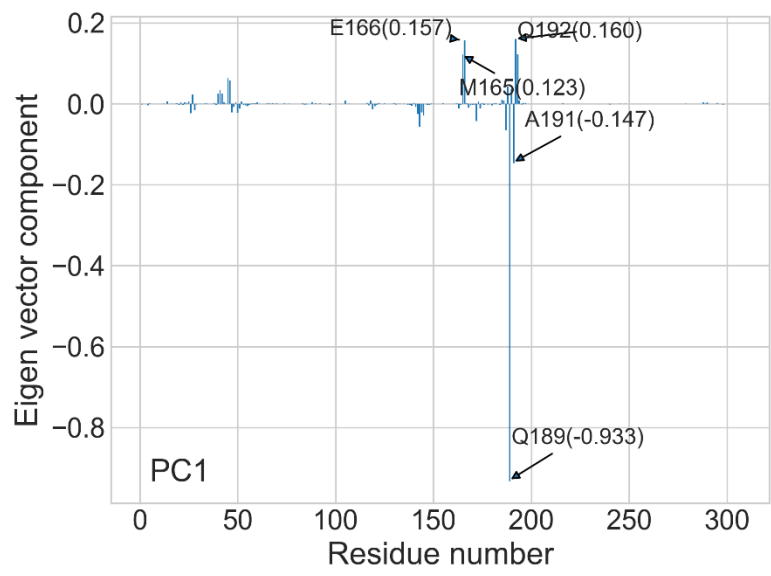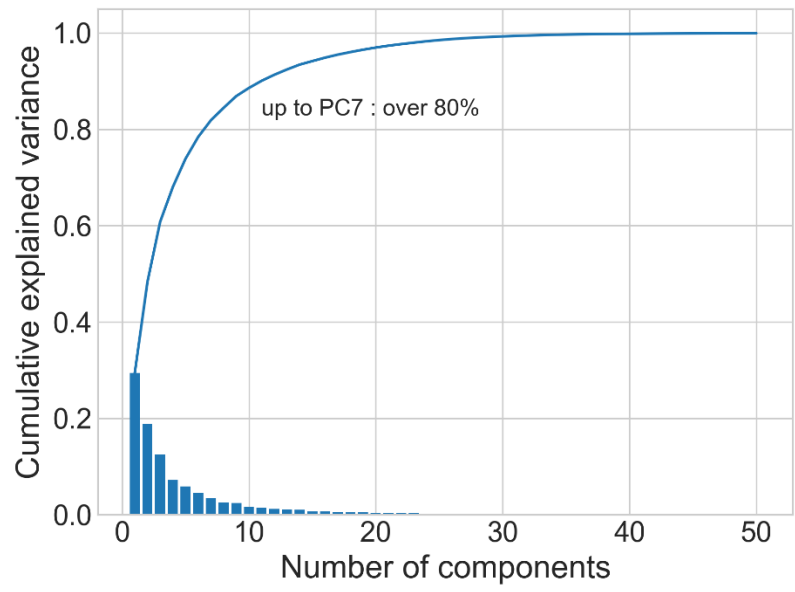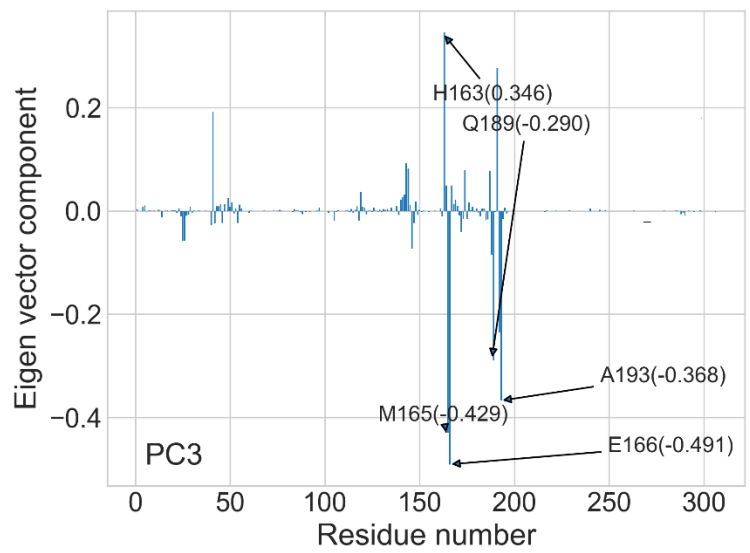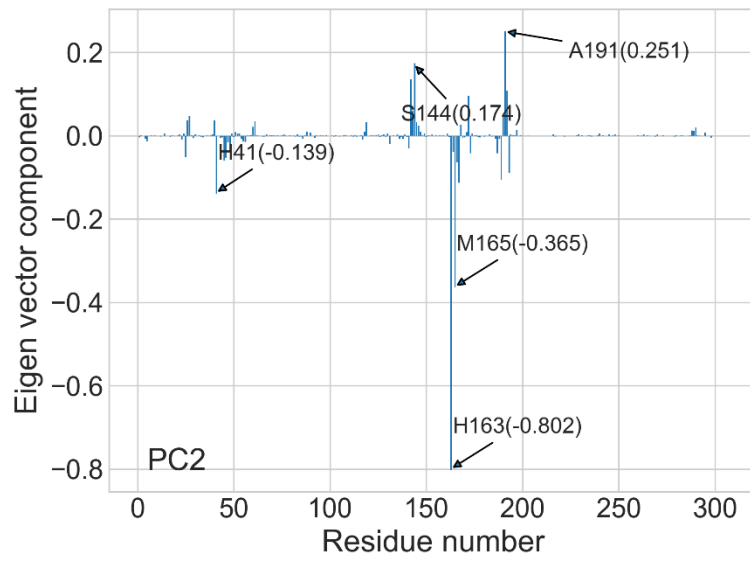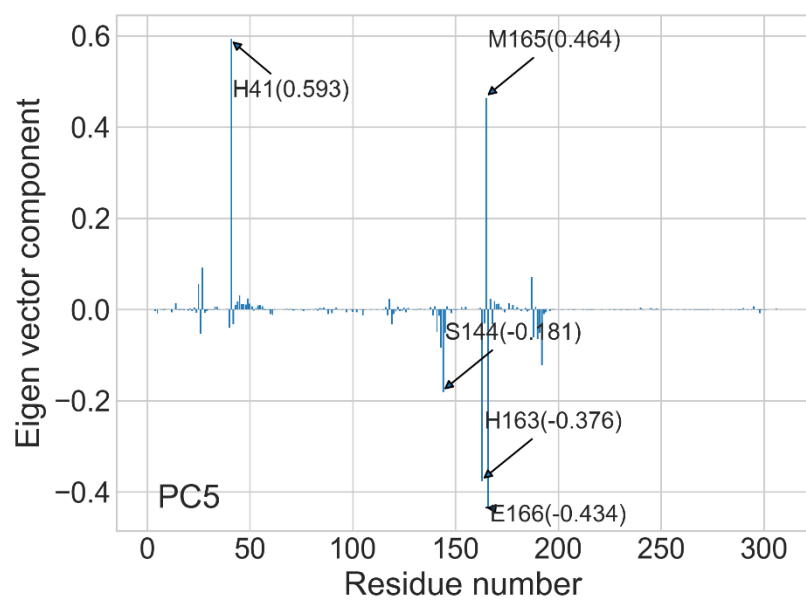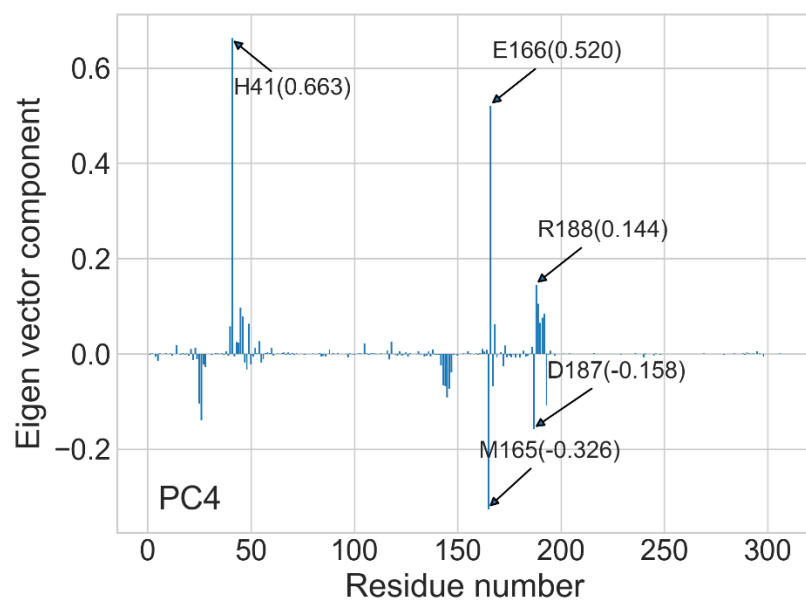
Figure S3. RMSDs for 100 ns MD in the absence of the covalent bond between N3 ligand and C145 residue of M^{pro}. Upper panel: C_α atoms in M^{pro}. Lower panel: C and N atoms in N3.

up to PC7 : over 80%

E166(0.157)
Q192(0.160)
M165(0.123)
A191(-0.147)
Q189(-0.933)
PC1

PC2

A191(0.251)

S144(0.174)

H41(-0.139)

M165(-0.365)

H163(-0.802)

PC3

H163(0.346)

Q189(-0.290)
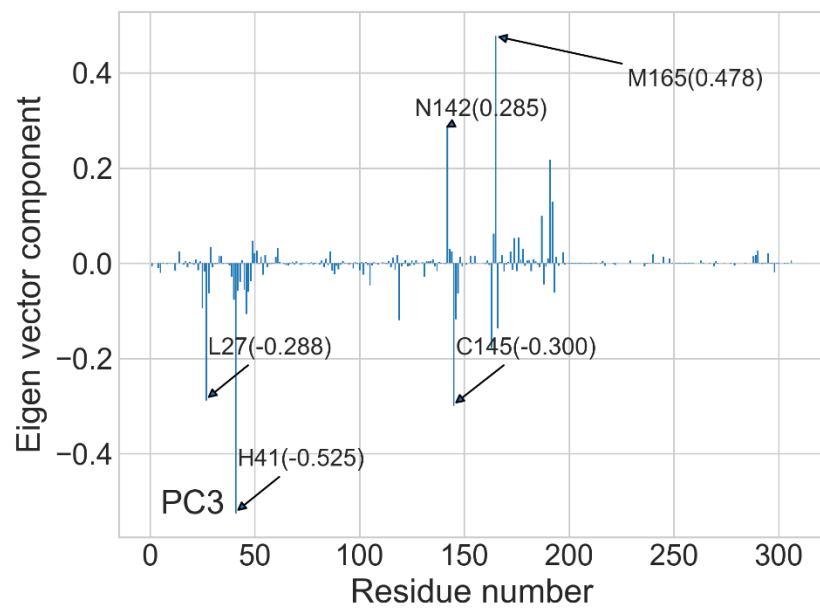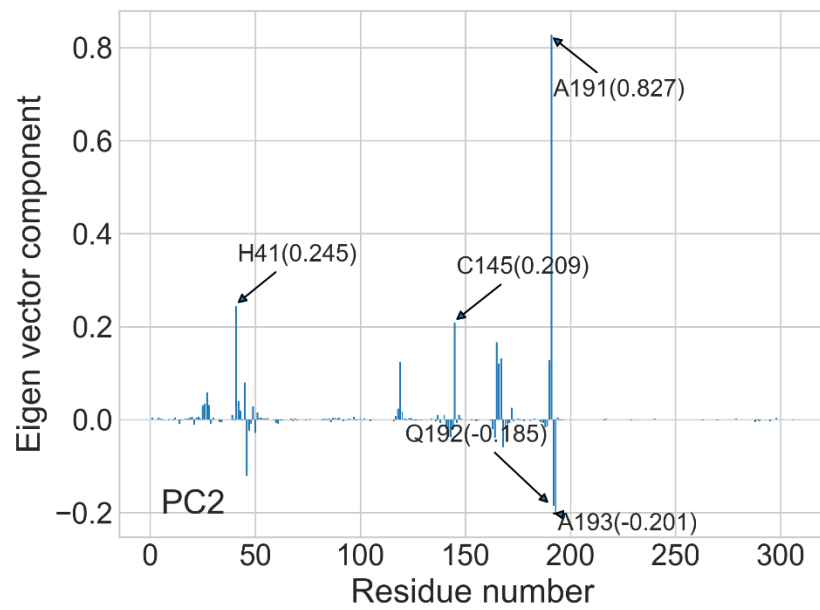
A193(-0.368)

M165(-0.429)

E166(-0.491)

Figure S4. PCA results for stage 1 (0-30 ns) in the case of the absence of the covalent bond between N3 and C145. Following the distribution of the eigenvalues, the residue components for the eigenvectors of PC1-PC5 are illustrated.
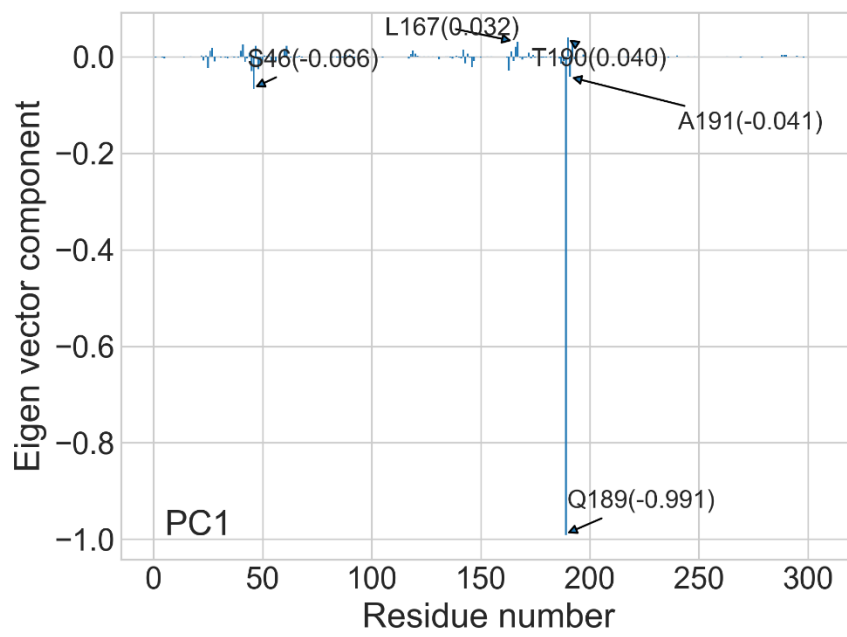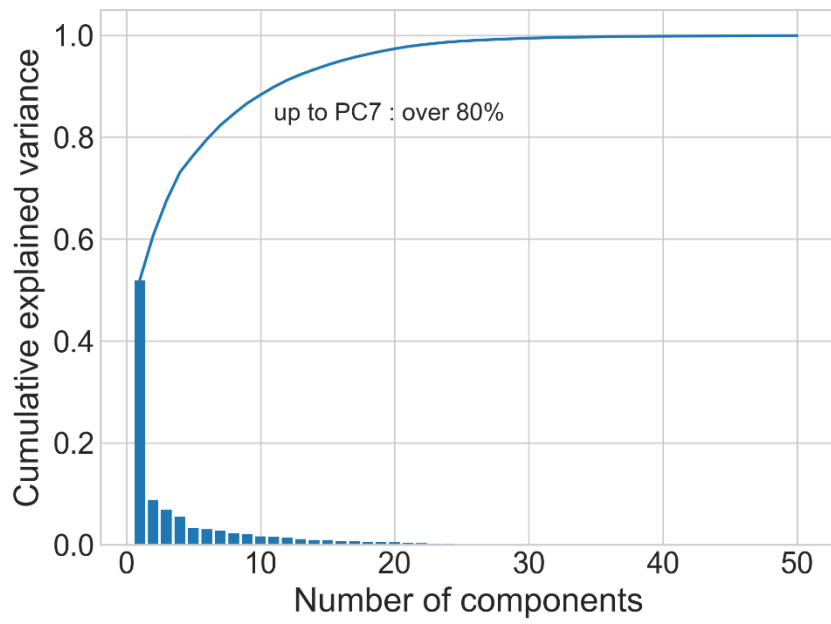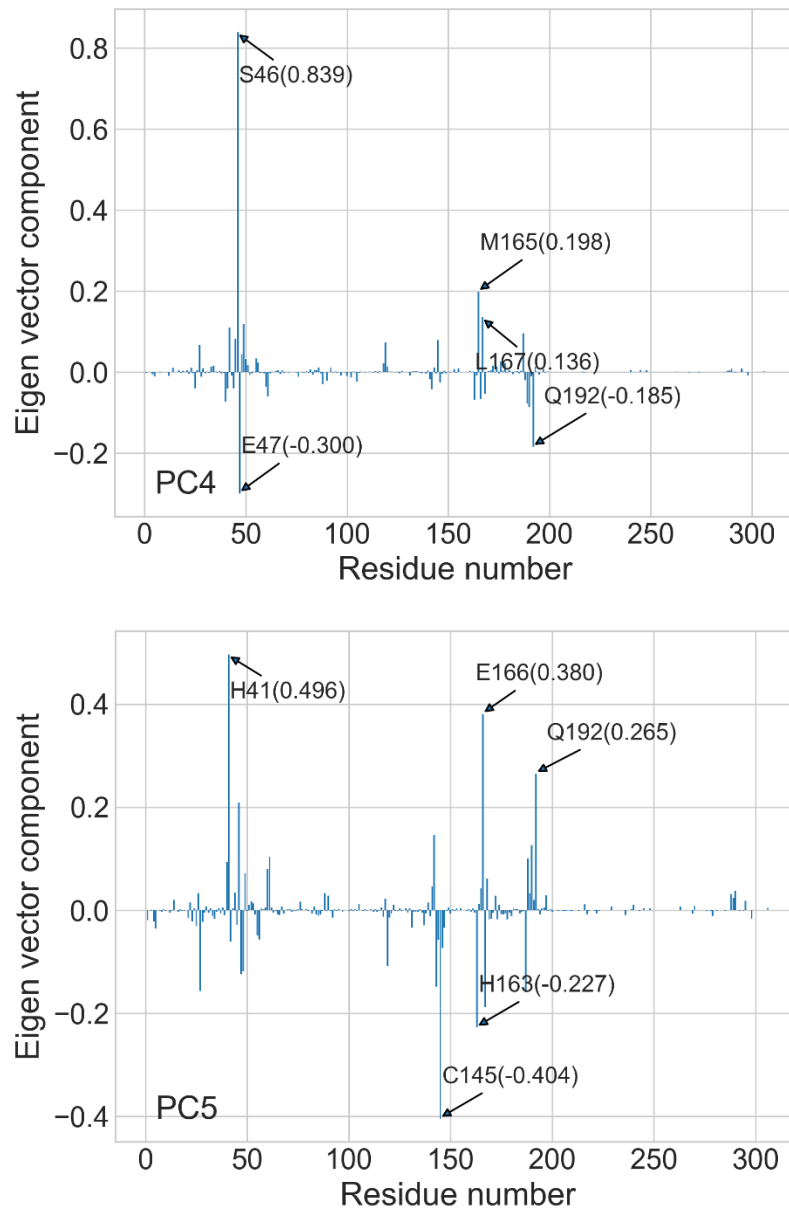
Figure S5. PCA results for stage 2 (30-100 ns) in the case of the absence of the covalent bond between N3 and C145. Following the distribution of the eigenvalues, the residue components for the eigenvectors of PC1-PC5 are illustrated.