

Published in final edited form as:

Proteins. 2008 November 15; 73(3): 742–753. doi:10.1002/prot.22094.

HIV-1 Protease Function and Structure Studies with the Simplicial Neighborhood Analysis of Protein Packing (SNAPP) Method

Shuxing Zhang¹, Andrew H. Kaplan², and Alexander Tropsha^{3,*}

¹Department of Experimental Therapeutics, M. D. Anderson Cancer Center, Unit 36, 1515 Holcombe Blvd, Houston, TX 77030

²Department of Microbiology and Immunology, UNC School of Medicine, Chapel Hill, NC 27599, USA

³The Laboratory for Molecular Modeling, Division of Medicinal Chemistry and Natural Products, UNC School of Pharmacy, Chapel Hill, NC 27599, USA

Abstract

The Simplicial Neighborhood Analysis of Protein Packing (SNAPP) method was used to predict the effect of mutagenesis on the enzymatic activity of the HIV-1 protease (HIVP). SNAPP relies on a four-body statistical scoring function derived from the analysis of spatially nearest neighbor residue compositional preferences in a diverse and representative subset of protein structures from the Protein Data Bank. The method was applied to the analysis of HIVP mutants with residue substitutions in the hydrophobic core as well as at the interface between the two protease monomers. Both wild type and tethered structures were employed in the calculations. We obtained a strong correlation, with R^2 as high as 0.96, between Δ SNAPP score (i.e., the difference in SNAPP scores between wild type and mutant proteins) and the protease catalytic activity for tethered structures. A weaker but significant correlation was also obtained for non-tethered structures as well. Our analysis identified residues both in the hydrophobic core and at the dimeric interface (DI) that are very important for the protease function. This study demonstrates a potential utility of the SNAPP method for rational design of mutagenesis studies and protein engineering.

Keywords

HIV-1 Protease (HIVP); Mutation; Tethered Dimer; Protein Packing; Delaunay Tessellation; Dimeric Interface (DI); Protein Stability; Catalytic Activity

INTRODUCTION

The human immunodeficiency virus type 1 (HIV-1) is an etiologic agent of the acquired immunodeficiency syndrome (AIDS). During virus assembly, the precursors that encode the structural proteins and enzymes that comprise the viral core are processed by a viral protease^{1,2}. This protease plays a critical role in viral replication. Mutations that inactivate the enzyme result in the production of non-infectious viral particles, and drugs that inhibit its activity have proven very useful in controlling HIV-associated clinical diseases^{2–4}.

As is the case for all retroviruses, the HIV-1 protease (HIVP) is an aspartic protease and is only functional as a dimer. The active enzyme consists of two identical subunits that form a dimer

*Corresponding author: Alexander Tropsha, Telephone (919) 966-2955, FAX: (919) 966-0204, alex_tropsha@unc.edu.

with a twofold (C2) symmetry⁵. Crystallographic studies indicate that the dimeric interface (DI) is composed of eight interacting N- and C-terminal residues (residues 1–4 and 96–99) of each chain. These residues are located within a four-stranded β -sheet (Figure 1), with polar side chains of Gln2, Thr4, Thr96 and Asn98 exposed to the solvent whereas hydrophobic residues Pro1, Ile3, Leu97 and Phe99 oriented towards the interior of the enzyme⁶. It was proposed that the four C-terminal residues (96–99) were particularly important in that they were involved in extensive inter-chain interactions to maintain the stability of the dimeric structure⁷, including 34 hydrogen bonds and 4 salt bridges⁸. Dimerization leads to the formation of an active site pocket consisting of the triads D25-T26-G27 from both monomers, and the two D25 residues are involved in the catalysis⁸.

Although the structure of the DI has been determined, individual contributions of interfacial residues to maintaining the protein stability and integrity have not been fully characterized. The eight N-terminal and C-terminal residues are conserved among different HIV-1 variants but there is considerable sequence variation between different retroviruses^{9,10}. An interesting observation that the large degree of sequence variation did not disrupt the four-stranded β -sheet structure and the protease activity was reported^{11,12}. This feature made the DI an attractive target for the design of novel therapeutic agents in order to inhibit the dimerization of those mutated, resistant enzymes^{13,14}.

In order to identify critical residues in the HIVP, numerous mutagenesis studies of the protease have been conducted^{4,15,16}. Herein, we present a computational analysis of the relationships between the structural stability and enzymatic activity of HIVP mutants using methodologies developed in our laboratory on the basis of a computational geometry approach known as Delaunay tessellation^{17–21}. This approach was introduced in the field of protein structure analysis by Finney and Richards^{22,23}. When applied to a collection of points in 3D space, the Delaunay tessellation partitions this structure into an aggregate of space-filling, irregular tetrahedra (or Delaunay simplices) with original points serving as vertices of these tetrahedra (Figure 2). To make the application of Delaunay tessellation to proteins, we use a reduced, united residue representation of a protein, in which each residue is represented by its side-chain centroid. With this set of united residues, Delaunay tessellation reduces the complex three-dimensional network of interactions in a protein to a collection of tertiary residue quadruplet motifs. Earlier, we developed a four-body fold recognition scoring function²⁴ and termed the approach Simplicial Neighborhood Analysis of Protein Packing (SNAPP)¹⁸ that uses Delaunay tessellation for the analysis of protein structure. The associated four-body score was called the SNAPP score.

Previously we successfully employed this approach for calculating the effects of hydrophobic core mutations on the stability of several proteins.¹⁸ A strong correlation was shown between the change of SNAPP scores as a result of mutations and the change of experimental thermodynamic stability with an $R = 0.86$.¹⁸ These results compared favorably with those in other studies ($R = 0.80$ or 0.79),^{25,26} which included many of the same mutations but employed distance-dependent potentials supplemented by local torsional potentials to estimate the $\Delta\Delta G_{\text{unfold}}$ values. Later, the SNAPP method was applied to study the relative stability of liganded vs. ligand-free protein conformations²⁰. We found that in most cases, the open form was more stable than the closed conformation, which was in agreement with available experimental observations²⁰. We have also demonstrated that SNAPP scores could discriminate between folding intermediates corresponding to pre- and post-transition states on the folding trajectory of the chymotrypsin inhibitor (CI2) protein and its native structure²⁷.

The major goal of the current study is to establish empirical models that predict the effect of mutations on the stability and activity of the HIVP. In many cases, the biological activity of a protein may be related to its thermodynamic stability. Therefore, we hypothesized that HIVP

catalytic activity may correlate with Δ SNAPP scores, which (as discussed above) were shown earlier to correlate with protein stability¹⁸. In addition, since resistant mutants frequently arise in patients treated with active site-directed inhibitors, novel drugs, e.g., targeting the DI, may prove more effective^{14,28–30}. Thus, we have also aimed to delineate the effect of individual amino acid and specific side chain interactions at the dimeric interface on the stability and activity of the protease.

METHODS

Experimental Data on HIVP Mutagenesis

Because the wild type functional enzyme is a homodimer, any mutation of the HIV-1 gene will affect both monomers identically. To facilitate studies of single mutations and understand the relative contribution of each half of the dimer, we produced a tethered system in which the C-terminal F99 of one subunit was joined to the N-terminal P1 of the other subunit by a flexible linker of 5 amino acids, G-T-S-S-G (Figure 1); this strategy was used earlier by several groups^{6,31,32}. It has been shown that the structure of a tethered dimer is nearly identical to the native, two-chain protease and has a wild-type substrate specificity.^{31,33} This construct enabled us to introduce mutations of interest into either half of the tethered structure in a single cloning step⁶.

In this study, experimental data on catalytic activity were analyzed for (I) various mutants of the native HIVP dimers (non-tethered) with double mutations; (II) various mutants of tethered homodimers with double and quadruple mutations; (III) various mutants of tethered heterodimers with single and double mutations. The experimental data was obtained as reported previously⁶. In brief, we modified an expression system that produced a single chain tethered structure of the protease; it was combined together with a heterologous β -galactosidase substrate altered by the insertion of HIVP cleavage site. The β -galactosidase substrate provided a convenient color screen, which enabled us to evaluate rapidly the phenotype of mutations introduced asymmetrically into either subunit of the dimer. In addition, by using a monoclonal antibody specific for β -galactosidase and quantitative Western blot, we were able to estimate the relative activity of these mutants. The expression of wild-type HIVP produced white colony and fully cleaved β -galactosidase whereas the expression of inactive enzymes such as D25N mutants produced blue colony and full-length β -galactosidase. Non-tethered system was used for comparison with tethered ones and to evaluate the impact of the linker on the protease activity. The experimental data used in this paper are given in Table 1.

SNAPP Analysis

SNAPP scores were calculated as described previously¹⁸. Briefly, the calculations were run as follows: (1) Create the tessellation pattern for the native protein. (2) Calculate the wild-type SNAPP score. The four-body SNAPP score was derived from non-redundant subsets of the PDB as a log-likelihood of occurrence of various quadruplet compositions of residues forming Delaunay simplices in tessellated protein structures. Thus, the SNAPP score, q , is derived as:

$q_{ijkl} = \log \frac{f_{ijkl}}{p_{ijkl}}$ where i, j, k, l are any four amino acid residues, f_{ijkl} is the frequency of occurrence of a given quadruplet as a Delaunay tetrahedron in the structural database, and p_{ijkl} is the expected frequency of occurrence of a given quadruplet based on individual amino acid frequencies in the same database. The overall SNAPP score is obtained as the sum of factors

q for all quadruplets of amino acids observed in a protein after tessellation, i.e., $S_{\text{SNAPP}} = \sum_{i=1}^{n_q} q_i$ where q_i is the statistical score for i -th quadruplet and n_q is the total number of Delaunay tetrahedra in the protein. (3) Change residue identity(-ies) to that of a mutant. Obviously, this

modification does not affect the structure of the protein so the mutant residues participate in the same tetrahedra as the wild-type residues. (4) Re-calculate the SNAPP score. The set of Delaunay simplices in the “mutant” model will be the same in terms of the tessellation pattern. However, the compositions of simplices will be different, resulting in different SNAPP scores. (5) Calculate $\Delta S_{\text{SNAPP}} = S_{\text{SNAPP}}(\text{mut}) - S_{\text{SNAPP}}(\text{wt})$. In practice, as indicated above, this difference is defined only by the new compositions of tetrahedra in which the mutated residue participates. Each residue was represented by the geometrical centroid of its side chain. There are many ways to define residue types but in this study we employed the standard 20-letter residue alphabet for 20 amino acid residue types.

The structure of the wild type tethered ligand-free HIVP is not available, but the tethered and non-tethered proteins are known to have similar activity^{31,32}. Therefore we created a tethered single chain protease model for our calculations based on the dimeric structure (pdb code: 3PHV) which was in an unliganded open form³⁴. A dummy Gly residue was used to “connect” two chains of the protein. However it was placed 40Å away from the residues it was artificially “bonded” to (P1 and F99) so that its addition would not affect the Delaunay tessellation pattern of the whole structure. A liganded tethered crystal structure 1HVC³¹ was available for SNAPP analysis. The SNAPP calculations were performed using MuSE and ProCAM modules of the SNAPP web server (<http://mmlsun4.pha.unc.edu/3dworkbench.html>) and some other internally developed programs.

Residue Mutation Scan and Comprehensive Mutational Profile Analysis

This analysis was performed as described by Vaisman and co-workers³⁵. Here we studied both monomer and dimer structures, and the dataset included $19 \times 198 = 3762$ mutants. The calculations included the following steps: (1) mutate every residue in the structure to the other 19 residue types and calculate the SNAPP score for each mutant; (2) calculate the SNAPP change by taking the difference between the total scores of the mutant and wild-type for all of the mutants in every position along the protein chain; (3) average the resulting 19 values of ΔS_{SNAPP} at each of the 99 (for the monomer) or 198 (for the dimer) positions. The complete list of all average SNAPP changes for every residue position in the protease was termed comprehensive mutational profile.

RESULTS

Correlation between ΔS_{SNAPP} Scores and Catalytic Activity

We have applied the SNAPP analysis to mutants of the HIVP included in Table 1. For the tethered homodimers, the correlation coefficient between the ΔS_{SNAPP} score and HIV-1 catalytic activity was $R^2=0.87$ (Figure 3). After excluding a quadruple mutant 3A4A/3A4A, the R^2 increased to 0.96. Similar results were obtained for tethered heterodimeric structures where R^2 was as high as 0.76 after excluding two outliers corresponding to double mutations in one chain (Figure 4). In contrast, for the non-tethered homodimers, the correlation between ΔS_{SNAPP} scores and HIVP activity was weak, with R^2 of only 0.45 (Figure 5).

SNAPP Profile Analysis

In order to analyze the individual residue contributions to the protein stability we generated the SNAPP profiles for three structures: single chain (3PHV), unliganded dimer (3PHV), and the tethered liganded protein-ligand complex (1HVC). Several interesting features were observed (Figure 6). For instance, some residues in the ligand-free dimer had very high SNAPP profile scores. Most of them were hydrophobic residues such as I15 (SNAPP = 10.159), L24 (SNAPP = 16.578), I64 (SNAPP = 13.313), I66 (SNAPP = 14.076), I85 (SNAPP = 12.369), L97 (SNAPP = 12.308), F99 (SNAPP = 10.771). These residues were located either in the hydrophobic core or at the DI of the HIVP. In contrast, the active site residues and those residues

forming the flap had very low SNAPP profile values. Comparing the profiles of the monomer vs. dimer, it is worth noting that upon dimerization there is a substantial change of SNAPP scores for residues both at the DI and in the active site (Figure 7). Analysis of the liganded (1HVC) vs. ligand-free (3PHV) structure showed that there was a large change of SNAPP scores for the flap region residues (M46-V56) while the SNAPP score change for the DI residues varied depending on their positions (Figure 8). Other residues did not exhibit great alteration of SNAPP scores.

Comprehensive Mutational Profile Analysis

Systematic mutational scan analysis showed that there were large SNAPP score fluctuations for residues located mostly in the regions of hydrophobic core, flaps, active site and DI (Figure 9). A strong inverse correlation was observed between the mean change of the protein SNAPP scores, when a residue was mutated to the other 19 residues, and the native SNAPP score of that residue, with R^2 as high as 0.84 (Figure 10). This result is consistent with the previous report by Vaisman and co-workers³⁵.

Interfacial Tessellation Analysis

The interfacial residues were defined by the means of Delaunay tessellation as follows. If a residue in one chain shared a tetrahedron with at least one residue from another chain, the residue was defined as interfacial residue and the respective tetrahedron was called interfacial tetrahedron. Figure 11 and Figure 13 focus on contacts of residues that form interfacial tetrahedra (yellow) between the two chains. For the ligand-free dimer, there were about 150 interfacial tetrahedra (Figure 11). In Figure 12, an upside down Eiffel Tower-like structure was formed by about 200 interfacial tetrahedra. Figure 13 highlights these tetrahedra in the context of the whole protein. The red and blue represents tetrahedra with high positive scores and negative scores, respectively. To demonstrate the importance of the ligand, a protein-ligand complex 1HVC was used to create the Delaunay tessellation with each residue represented by its side chain centroid. The ligand is shown as a space filling model in Figure 13.

DISCUSSION

The residue contacts in protein structures could be represented as sets of two-body or multi-body interactions. The majority of statistical potentials employed to analyze protein structures use two-body terms³⁶. However, protein structures are complex residue interaction networks. The interactions between two atoms (or residues) are certainly affected by the presence of the third, fourth or more atoms (or residues). Therefore, multi-body interaction scoring functions appear natural and desirable. In addition, Delaunay tessellation in 3D space can unambiguously define quadruplets of nearest neighbors, and four-body contacts or 3D simplices are intrinsic to three dimensional nearest-neighbor patterns. In fact, we have demonstrated earlier that the four-body SNAPP scoring function afforded better discrimination between native structures and structural decoys than a popular two-body Miyazawa-Jernigan potential²⁷. Thus, four-body SNAPP scoring function was employed in this study.

The SNAPP statistical scoring function used in our calculations was derived from a set of 1922 protein structures (<2.5Å resolution, and <20% identity)³⁷. The proteins in this set are highly diverse, although some remote homologues are likely to be included. Thus, the SNAPP potential has no bias towards a particular protein family or fold that could affect the results of our calculations.

SNAPP Profile Analysis

In the unliganded HIVP dimeric structure, a repeated SNAPP residue profile pattern (residues 1–99 and 100–198) was observed as expected (Figure 6, 3PHV dimeric structure), which is

consistent with the fact that there is a twofold axis of symmetry traversing through the substrate-binding pocket³⁴. The residue profile analysis helped identify the important residues. As is well known, the hydrophobic core is critical for the stability of protein structure and consequently, for its function. Figure 6 shows the SNAPP profile for all residues in the HIVP. The scores reflect the environment of each residue and reveal the contribution of each residue to the stability of the entire protein. Our previous studies indicated that the higher the SNAPP score of a residue, the more deeply it was buried inside the hydrophobic environment^{18,37}. If the SNAPP score is below zero, the residue is likely to be on the surface and solvent accessible.

The gray dashed line in Figure 6 corresponds to an arbitrary cutoff value (five SNAPP units) that was used in the analysis. If the SNAPP score was higher than this value we considered the residue buried (and therefore important for protein stability)^{18,37}. For example, L24 has the highest value of the SNAPP score (Figure 6) and it is one of the hydrophobic core residues. The tessellation analysis (Figure 14) showed that this residue not only interacted with the active site (L23, D25, T26) and hydrophobic core (L90, I66, I85, etc.) residues, but also with those at the DI (I3, L97, F99, etc.). In total, L24 has 14 residues as its nearest neighbors. Although there is no experimental data suggesting that L24 interacts directly with the catalytic residue D25 or with protease inhibitors, we believe that it may serve as a bridge mediating the intra-molecular communications between the DI and the active site. Mutation of this residue to any of the other 19 residue types is predicted to dramatically decrease the structure stability. It is of note that Weber and co-workers recently demonstrated that L24I mutation led to an indinavir-resistant enzyme with the reduction of both stability and cleavage activity³⁸. Consistent with this observation, SNAPP analysis indicated that the score of this residue decreased 0.5 units for the L24I mutation.

The analysis of the HIVP structure indicated that residues P1, I3, H69, I93, C95, and L24 formed a hydrophobic pocket surrounding F99. This pocket could be viewed as a target site for possible design of a ligand capable of interfering with the dimerization. On the other hand, D25 and the flap residues showed low SNAPP scores, indicating that they were flexible and solvent-exposed, in agreement with molecular dynamics studies demonstrating that the flap underwent large local conformational changes upon ligand binding or release^{39,40}. D25 only formed two unfavorable tetrahedra (with residues L32, V82, I84, A28 and G86) possessing negative SNAPP scores. One could hypothesize that the relatively unstable local structure formed by these residues in the apo-protein will be stabilized by their interactions with the ligand in the bound complex.

The SNAPP residue differential profile analysis for the monomer vs. dimer (Figure 7) showed that several head and tail residues, such as residues 3, 5, 97 and 99, had large SNAPP changes upon dimerization. This is not surprising because it is well known that these residues play an important role in the dimer formation; they are involved in a plethora of interactions and stabilize the dimeric structure^{7,41}. Of note, several groups have designed DI peptidomimetic analogues that are effective in blocking enzyme dimerization^{13,14,28,29}. The mutation of these residues may change the stability of the dimer and thus lead to the alteration of the enzymatic activity, as demonstrated in this paper and in our earlier report⁶.

Residue L90, which is close to the DI, was also observed to exhibit a large SNAPP score change upon dimerization (Figure 7), indicating its significance in stabilizing the DI. X-ray analysis revealed that L90 interacted with the active site residue D25⁴². The tessellation also showed that L90 formed tetrahedra with residues around the active site (I24, T26, D25, etc.) and DI (L5, L97, F99, etc.). A recent antiretroviral resistance mutation study with 306 patients demonstrated that 41% of the resistant mutations occurred at L90⁴³. These important observations lend additional support to our hypothesis that SNAPP analysis could pinpoint individual residues responsible for the protein stability and functions.

Comparison of tethered ligand-free and tethered liganded dimers (Figure 8) showed that the conformational change led to large SNAPP score changes for DI residues, indicating that ligand binding induces conformational changes that translate towards and perturb the DI. Studies comparing available liganded and unliganded structures suggested that the core domain rotated approximately 7 degrees relative to the terminal domain, which forms the dimeric interface⁴⁴. Therefore, the interface plays a critical role in both forming the active site binding pocket and in accommodating the changes associated with the substrate binding. The mutation of these residues (that experience the significant change in the SNAPP score upon ligand binding) would most probably affect the hydrophobic core and thus influence the stability as well as the activity of the enzyme, making these residues good targets for experimental mutagenesis.

The data discussed above for residues L24 and L90 may suggest that both residues with high SNAPP scores for the monomer and those exhibiting large SNAPP profile change upon the dimerization may be responsible for inhibitor-induced resistance, which presents a testable hypothesis. From the SNAPP residue differential profile analysis (Figure 8) we also conclude that residues in the flap region and active site were flexible and undertook significant conformational changes upon ligand binding.

Comprehensive Mutational Profile Analysis

The systematic mutational scan analysis also helped us identify residues that were important for the enzyme stability and activity. Our results (Figure 9) demonstrated that the large fluctuation occurred mostly in the regions of the hydrophobic cores, flap, active site and the DI, consistent with our SNAPP residue profile analysis. The high fluctuation indicated that the structure was not stable as a result of mutations to other residue types. We also found that there was a strong inverse correlation between the mean mutational profile values and the native residue profile values ($R^2 = 0.84$, Figure 10). This suggested that the influence of mutations was highly dependent on the environment of the residues. Thus, residues with high SNAPP scores are clearly important for maintaining the structure stability, and their mutations to other residues are likely to destabilize the structure (and thus probably diminish the activity). It also implies that the highest scoring residues are the least resistant to any mutation. As a control, we shuffled the sequences of HIVP and repeated the above differential profile calculations, but no correlation was obtained. These studies indicated that the high inverse correlation between the mean mutational profile value and the native residue profile could be a feature unique to the native or near-native structures.

Correlation between Δ SNAPP Scores and HIVP Activity

Although we established earlier that the Δ SNAPP scores were highly correlated with protein thermodynamic stability¹⁸, in general there is no simple relationship between the stability and enzymatic activity. Nevertheless, as shown above, in some cases, we were able to confirm that for HIVP and its mutants, the activity and stability were highly correlated, especially for the DI mutations in the tethered structures. In Figure 3, the initial R^2 of the tethered homodimers was 0.50. However, after the active site mutant was excluded, the R^2 increased to 0.88. One possible explanation is that residue D25 formed critical interactions (e.g., hydrogen bonds) with the substrate; when mutations occur, such favorable interactions disappear. It is also possible that D25 is unique in forming the enzymatic transition state of the protease-ligand complex during catalysis. Due to obvious limitations of our method and simplified representation of the protein structure, we could not include the ligand in our calculations, and the dynamic catalytic process was not considered either, so we could not predict the effect of that mutation. Furthermore, when mutant 3A4A/3A4A was excluded, the correlation coefficient R^2 increased to 0.96. This quadruplet mutation could possibly induce either the allosteric structural change of the active site or simply disrupt the dimeric structure, which was

not captured in our calculations (that are restricted to analyzing only the available static structures).

These results are significant for several reasons: (1) with the (very simple) SNAPP method, we could actually predict activity for certain HIVP mutants using the model constructed here; (2) the mutations at the interface probably led to the disruption or misfolding of the non-tethered dimeric structure, which made them unstable and consequently inactive; and (3) the tethered systems were more stable than the non-tethered systems and this was probably due to the function of the linker that could keep the two monomers from getting away upon structural perturbations caused by mutations.

Although our results showed that the HIVP activity correlated well with Δ SNAPP scores for tethered systems, the correlation was weak in the case of non-tethered system (Table I and Figure 5). There was no obvious trend in the data even after excluding structures with multiple mutations. One possible reason is that in tethered systems, the protease could, to some extent, resist structural perturbations that are caused by mutations. The linker probably could keep the two monomers together and make the structure more stable. In contrast, the non-tethered systems were more labile. The mutations perturbed the interfacial interactions thereby leading to significant structural change or even the disruption of the dimer, which can not be captured in the SNAPP analysis. With the tethered heterodimers, this hypothesis proved to be valid as we observed a fairly good correlation between Δ SNAPP score and enzymatic activity (Figure 4): the original R^2 was 0.52 and it increased to 0.76 after excluding two outliers that corresponded to double mutations (3A4A/WT and 97A98A/WT). This again indicated that the multiple mutations could cause conformational change but the effect could not be captured by our model. Another explanation of these observations is that our SNAPP method, which was developed based on four-body contacts within monomeric proteins, is more appropriate for tethered systems which are more similar to monomeric structures.

Interfacial Tessellation Analysis

Interfacial tessellation study revealed that around 200 tetrahedra were formed between the two monomers in the liganded structure. Obviously, these interfacial interactions are responsible for stabilizing the dimeric structure. The mutation of interfacial residues included in these (especially high scoring) tetrahedra is likely to affect the stability and activity of the protease. The interfacial tetrahedra (yellow, red and blue) are shown in Figure 11 and Figure 12 for unliganded (3PHV) and liganded (1HVC) HIVP, respectively. The most obvious difference of the tessellation patterns between liganded and unliganded structures is the formation of additional tetrahedra in the liganded complex around the active site, flap region, and the four-stranded β -sheet structure at the DI that were not observed in the unliganded dimer.

Figure 15 shows a high scoring tetrahedron, which includes F199 from one chain (labeled as PHE199) and three residues from the other chain. L24 has a very high SNAPP profile score and forms interactions with many residues as described previously. C95 and F199 are at the DI, and I93 is in the hydrophobic core. This data suggests that mutations of these important residues are likely to influence the dimer conformation or the activity, or both. A better understanding of the role of these residues in protein stability and activity could help developing new therapeutics. Very interestingly, Hosur and co-workers indeed found earlier by X-ray analysis at 1.9Å resolution that the unliganded tethered structure with C95M mutation had a closed flap conformation⁴⁵. In addition, C95 has been targeted for the design of a novel HIVP inhibitor that was shown to have activity at 3.7μM¹⁴.

Similarly, another set of four residues, N98, N198, T196 and Q2 (not shown), are located at the dimeric interface and the tetrahedron that is formed by these residues has a high SNAPP score. Since only the dimerized protease has the cleavage activity, it is reasonable to

hypothesize that these residues that form the interfacial high-scoring tetrahedra play an important role in both stabilizing the dimer and affecting its activity. Characterizing these residues would be very helpful in the study of HIVP structure and function. In particular, residues L3, L5, P9, L24, T26, G49, I50, G51 and C90-F99 were found to have the highest total number of interfacial contacts as revealed by the Delaunay tessellation analysis and they were pursued in the experimental mutagenesis studies in one of our laboratories.

In addition to residues that form high-scoring tetrahedra, those with negative scores could be critical for protein function as well. Kapustina et al. recently employed SNAPP to analyze the allosteric behavior of the tryptophanyl-tRNA synthetase⁴⁶. They found that some negatively scored tetrahedral motifs might play an important role in determining the relative stability of different allosteric conformers. Enlightened by that study, we conducted the analysis of the distribution of high-scoring and negative-scoring tetrahedra⁴⁶. As indicated in Figure 11–Figure 13, this analysis illustrated that interactions at the dimeric interface (red tetrahedra) stabilize both unliganded and liganded structures. Interestingly, we found that many of the negatively scored tetrahedra (blue tetrahedra) were close to the active site, and that those residues included L24/L124, D25/D125, G27/G127, etc. Figure 13 also indicates that some newly formed tetrahedra around the flap region in liganded complex are not stable (i.e., have negative scores). This observation demonstrated that these unfavorable tetrahedra are important for the function of HIV protease in that they allow the flexibility of the structure and thus local conformation can be changed to accommodate the substrate binding and release.

Tetrahedra formed at the interface between the two monomer chains were defined as interfacial tetrahedra by the Delaunay tessellation. It is presently unknown whether or not the compositional requirements to the interfacial tetrahedra differ from those already defined for the protein interior tetrahedra. Thus, we currently rely on the existing SNAPP score developed in our laboratory earlier for the single chain proteins. We are also developing a separate scoring function specific for the protein-protein interface derived from the tessellation analysis of various protein-protein complexes available in the PDB. Preliminary data (not shown) suggest that the SNAPP scores derived from the analysis of protein-protein interfaces are similar to those derived from internal tetrahedra. This ongoing project will help us delineate both general and unique requirements to the composition and geometry of amino acid residues located at the interface of protein-protein complexes, as compared to those we have already established for individual protein chains.

CONCLUSIONS

This study was based on our previous successful experience in calculating the effect of residue mutations on protein stability using the simple SNAPP scoring function. We have generally assumed that the enzyme activity and stability are related. The result here, to our knowledge, is the first attempt to demonstrate that under certain conditions (e.g., no large structure change), the correlation between activity and a statistical scoring function that reflects protein stability can indeed be established. The correlation is particularly good for tethered mutant systems with R^2 as high as 0.96.

In addition, the SNAPP and residue profile analysis helped us identify individual residues important for the stability and activity, which were not immediately apparent from crystallographic studies of the protein structure. Furthermore, the interfacial tessellation analysis demonstrated that those interfacial residues formed strong interactions to maintain the dimer stability. We have also shown that residues identified as important for maintaining protease stability, either in the hydrophobic core or at the interface, may form a network of contacts communicating the perturbations occurring either at the interface or in the active site across the residue interaction network. We hypothesized that such residues could be

responsible for HIVP resistance to current medications and cited several experimental papers that lend support for this hypothesis. We shall continue to use the SNAPP method to predict the activity change due to various mutations, or to locate residues that appear to control the stability, activity, and drug resistance. These predictions provide research hypotheses that can be tested experimentally. Further development of a more robust algorithm for these calculations is in progress and we shall rely on the interplay between experimental and computational research to help us refine and improve this approach.

ACKNOWLEDGEMENTS

This paper is dedicated to the memory of Dr. Andrew Kaplan who devoted himself to HIV studies, and who initiated this project in 2001. Andy tragically passed away unexpectedly in the midst of his career at the age of 47 in 2006, and the co-authors regard this paper as a special memoriam for a very special colleague. We thank Dr. Jun Feng for his help with the interfacial tessellation analysis. This work was supported by the NIH research grants GM068665 (AT) and GM066681 (AK).

REFERENCES

1. Briggs JA, Grunewald K, Glass B, Forster F, Krausslich HG, Fuller SD. The Mechanism of HIV-1 Core Assembly: Insights from Three-Dimensional Reconstructions of Authentic Virions. *Structure* 2006;14:15–20. [PubMed: 16407061]
2. Kaplan AH. Assembly of the HIV-1 core particle. *AIDS Rev* 2002;4:104–111. [PubMed: 12152518]
3. Kaplan AH, Scheyett A, Golin CE. HIV and stigma: analysis and research program. *Curr HIV /AIDS Rep* 2005;2:184–188. [PubMed: 16343376]
4. Shafer RW, Rhee SY, Pillay D, Miller V, Sandstrom P, Schapiro JM, Kuritzkes DR, Bennett D. HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *AIDS* 2007;21:215–223. [PubMed: 17197813]
5. Katoh I, Yasunaga T, Ikawa Y, Yoshinaka Y. Inhibition of retroviral protease activity by an aspartyl proteinase inhibitor. *Nature* 1987;329:654–656. [PubMed: 2821409]
6. Choudhury S, Everitt L, Pettit SC, Kaplan AH. Mutagenesis of the dimer interface residues of tethered and untethered HIV-1 protease result in differential activity and suggest multiple mechanisms of compensation. *Virology* 2003;307:204–212. [PubMed: 12667791]
7. Weber IT. Comparison of the crystal structures and intersubunit interactions of human immunodeficiency and Rous sarcoma virus proteases. *J Biol Chem* 1990;265:10492–10496. [PubMed: 2162350]
8. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J, Kent SB. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* 1989;245:616–621. [PubMed: 2548279]
9. Blundell TL, Cooper JB, Sali A, Zhu ZY. Comparisons of the sequences, 3-D structures and mechanisms of pepsin-like and retroviral aspartic proteinases. *Adv Exp Med Biol* 1991;306:443–453. [PubMed: 1812741]
10. Weber IT. Structural alignment of retroviral protease sequences. *Gene* 1989;85:565–566. [PubMed: 2560758]
11. Gustchina A, Kervinen J, Powell DJ, Zdanov A, Kay J, Wlodawer A. Structure of equine infectious anemia virus proteinase complexed with an inhibitor. *Protein Sci* 1996;5:1453–1465. [PubMed: 8844837]
12. Vondrasek J, van Buskirk CP, Wlodawer A. Database of three-dimensional structures of HIV proteinases. *Nat Struct Biol* 1997;4:8. [PubMed: 8989313]
13. Shultz MD, Chmielewski J. Probing the role of interfacial residues in a dimerization inhibitor of HIV-1 protease. *Bioorg Med Chem Lett* 1999;9:2431–2436. [PubMed: 10476882]
14. Zutshi R, Chmielewski J. Targeting the dimerization interface for irreversible inhibition of HIV-1 protease. *Bioorg Med Chem Lett* 2000;10:1901–1903. [PubMed: 10987413]
15. Ravela J, Betts BJ, Brun-Vezinet F, Vandamme AM, Descamps D, van LK, Smith K, Schapiro JM, Winslow DL, Reid C, Shafer RW. HIV-1 protease and reverse transcriptase mutation patterns

- responsible for discordances between genotypic drug resistance interpretation algorithms. *J Acquir Immune Defic Syndr* 2003;33:8–14. [PubMed: 12792349]
16. Skalova T, Dohnalek J, Duskova J, Petrokova H, Hradilek M, Soucek M, Konvalinka J, Hasek J. HIV-1 protease mutations and inhibitor modifications monitored on a series of complexes. Structural basis for the effect of the A71V mutation on the active site. *J Med Chem* 2006;49:5777–5784. [PubMed: 16970402]
 17. Singh RK, Tropsha A, Vaisman II. Delaunay tessellation of proteins: four body nearest-neighbor propensities of amino acid residues. *J Comput Biol* 1996;3:213–221. [PubMed: 8811483]
 18. Carter CW Jr, LeFebvre BC, Cammer SA, Tropsha A, Edgell MH. Four-body potentials reveal protein-specific correlations to stability changes caused by hydrophobic core mutations. *J Mol Biol* 2001;311:625–638. [PubMed: 11518520]
 19. Oloff S, Zhang S, Sukumar N, Breneman C, Tropsha A. Chemometric analysis of ligand receptor complementarity: identifying Complementary Ligands Based on Receptor Information (CoLiBRI). *J Chem Inf Model* 2006;46:844–851. [PubMed: 16563016]
 20. Sherman DB, Zhang S, Pitner JB, Tropsha A. Evaluation of the relative stability of liganded versus ligand-free protein conformations using Simplicial Neighborhood Analysis of Protein Packing (SNAPP) method. *Proteins* 2004;56:828–838. [PubMed: 15281134]
 21. Zhang S, Golbraikh A, Tropsha A. Development of quantitative structure-binding affinity relationship models based on novel geometrical chemical descriptors of the protein-ligand interfaces. *J Med Chem* 2006;49:2713–2724. [PubMed: 16640331]
 22. Finney, JL. Proceedings of the Royal Society. 1970. Voronoi; p. 479–493.
 23. Richards FM. The interpretation of protein structures: total volume, group volume distributions and packing density. *J Mol Biol* 1974;82:1–14. [PubMed: 4818482]
 24. Zheng W, Cho SJ, Vaisman II, Tropsha A. A new approach to protein fold recognition based on Delaunay tessellation of protein structure. *Pac Symp Biocomput* 1997:486–497. [PubMed: 9390317]
 25. Gilis D, Rooman M. Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local versus non-local interactions along the sequence. *J Mol Biol* 1997;272:276–290. [PubMed: 9299354]
 26. Topham CM, Srinivasan N, Blundell TL. Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. *Protein Eng* 1997;10:7–21. [PubMed: 9051729]
 27. Krishnamoorthy B, Tropsha A. Development of a four-body statistical pseudo-potential to discriminate native from non-native protein conformations. *Bioinformatics* 2003;19:1540–1548. [PubMed: 12912835]
 28. Boggetto N, Reboud-Ravaux M. Dimerization inhibitors of HIV-1 protease. *Biol Chem* 2002;383:1321–1324. [PubMed: 12437124]
 29. Pettit SC, Gulnik S, Everitt L, Kaplan AH. The dimer interfaces of protease and extra-protease domains influence the activation of protease and the specificity of GagPol cleavage. *J Virol* 2003;77:366–374. [PubMed: 12477841]
 30. Frutos S, Rodriguez-Mias RA, Madurga S, Collinet B, Reboud-Ravaux M, Ludevid D, Giralt E. Disruption of the HIV-1 protease dimer with interface peptides: structural studies using NMR spectroscopy combined with [2-(13)C]-Trp selective labeling. *Biopolymers* 2007;88:164–173. [PubMed: 17236209]
 31. Bhat TN, Baldwin ET, Liu B, Cheng YS, Erickson JW. Crystal structure of a tethered dimer of HIV-1 proteinase complexed with an inhibitor. *Nat Struct Biol* 1994;1:552–556. [PubMed: 7664084]
 32. Cheng YS, Yin FH, Foundling S, Blomstrom D, Kettner CA. Stability and activity of human immunodeficiency virus protease: comparison of the natural dimer with a homologous, single-chain tethered dimer. *Proc Natl Acad Sci U S A* 1990;87:9660–9664. [PubMed: 2263618]
 33. Phylip LH, Griffiths JT, Mills JS, Graves MC, Dunn BM, Kay J. Activities of precursor and tethered dimer forms of HIV proteinase. *Adv Exp Med Biol* 1995;362:467–472. [PubMed: 8540359]
 34. Lapatto R, Blundell T, Hemmings A, Overington J, Wilderspin A, Wood S, Merson JR, Whittle PJ, Danley DE, Geoghegan KF. X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature* 1989;342:299–302. [PubMed: 2682266]

35. Masso M, Vaisman II. Comprehensive mutagenesis of HIV-1 protease: a computational geometry approach. *Biochem Biophys Res Commun* 2003;305:322–326. [PubMed: 12745077]
36. Lu H, Skolnick J. A distance-dependent atomic knowledge-based potential for improved protein structure selection. *Proteins* 2001;44:223–232. [PubMed: 11455595]
37. Tropsha A, Carter CW Jr, Cammer SA, Vaisman II. A Computational Geometry Approach to Studying Proteins. *Methods in Enzymol* 2003;374:509–544. [PubMed: 14696387]
38. Liu F, Boross PI, Wang YF, Tozser J, Louis JM, Harrison RW, Weber IT. Kinetic, stability, and structural changes in high-resolution crystal structures of HIV-1 protease with drug-resistant mutations L24I, I50V, and G73S. *J Mol Biol* 2005;354:789–800. [PubMed: 16277992]
39. Collins JR, Burt SK, Erickson JW. Flap opening in HIV-1 protease simulated by 'activated' molecular dynamics. *Nat Struct Biol* 1995;2:334–338. [PubMed: 7796268]
40. Hornak V, Simmerling C. Targeting structural flexibility in HIV-1 protease inhibitor binding. *Drug Discov Today* 2007;12:132–138. [PubMed: 17275733]
41. Wlodawer A, Vondrasek J. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu Rev Biophys Biomol Struct* 1998;27:249–284. [PubMed: 9646869]
42. Hong L, Zhang XC, Hartsuck JA, Tang J. Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: insights into the mechanisms of drug resistance. *Protein Sci* 2000;9:1898–1904. [PubMed: 11106162]
43. Rodrigues R, Vazquez CM, Colares JK, Custodio RM, Bonasser FF, Souza LR, Gianna MC, Marques CC, Brígido LF. Antiretroviral resistance mutations in human immunodeficiency virus type 1 infected patients enrolled in genotype testing at the Central Public Health Laboratory, Sao Paulo, Brazil: preliminary results. *Mem Inst Oswaldo Cruz* 2005;100:97–102. [PubMed: 15867972]
44. Rose RB, Rose JR, Salto R, Craik CS, Stroud RM. Structure of the protease from simian immunodeficiency virus: complex with an irreversible nonpeptide inhibitor. *Biochemistry* 1993;32:12498–12507. [PubMed: 8241141]
45. Pillai B, Kannan KK, Hosur MV. 1.9 Å x-ray study shows closed flap conformation in crystals of tethered HIV-1 PR. *Proteins* 2001;43:57–64. [PubMed: 11170214]
46. Kapustina M, Weinreb V, Li L, Kuhlman B, Carter CW Jr. A conformational transition state accompanies tryptophan activation by *B. stearothermophilus* tryptophanyl-tRNA synthetase. *Structure* 2007;15:1272–1284. [PubMed: 17937916]

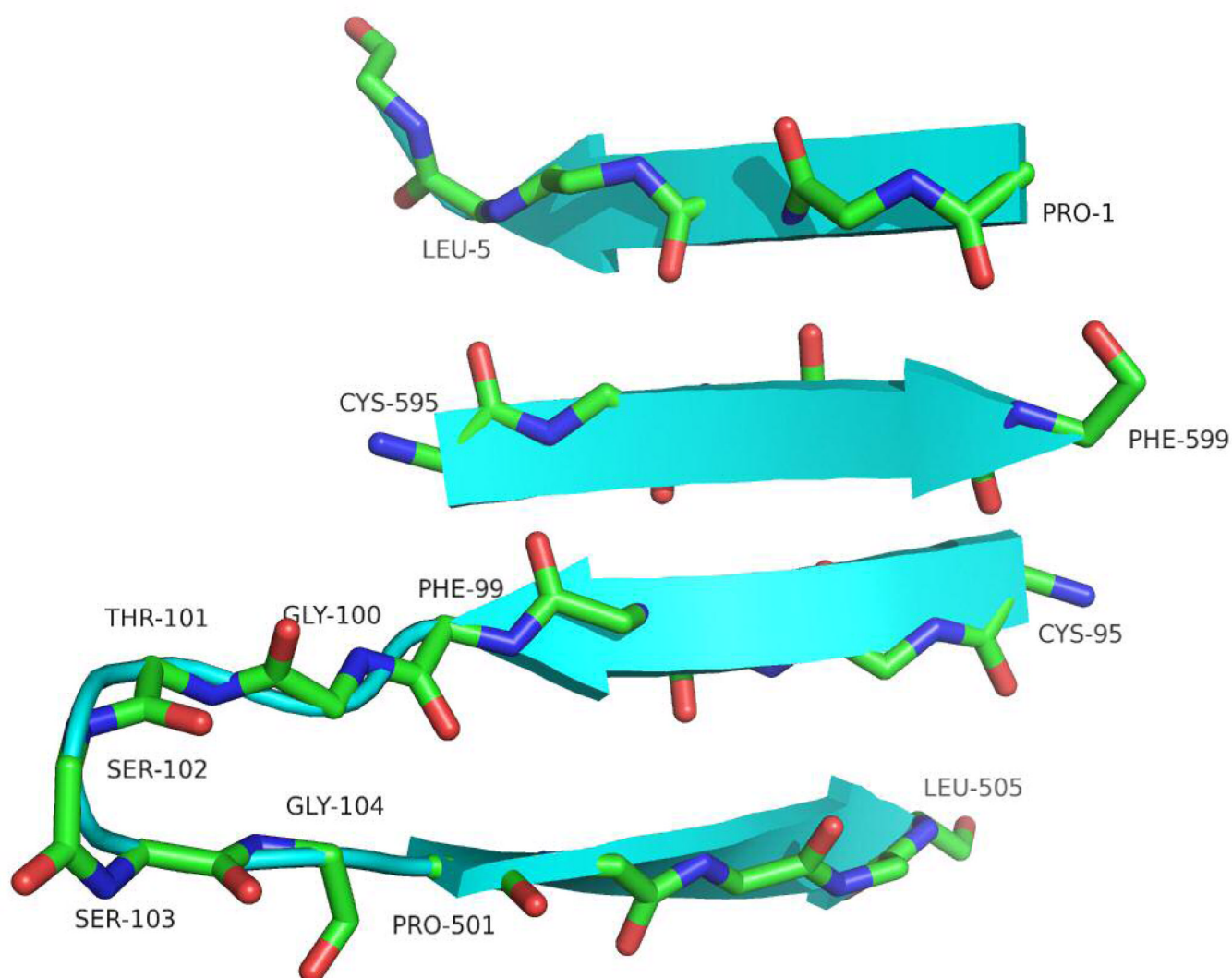


Figure 1.

The head and tail residues of the tethered HIV-1 protease (only the backbone is shown) form four stranded β -sheets (cyan). The linker G-T-S-S-G (residue number 100 ~ 104) connects the two subunits. The structure was created based on 1HVC with SYBYL (Tripos, Inc., St. Louis, MO) and visualized with PyMol (DeLano Scientific LLC, San Francisco, CA). 42×34mm (600 × 600 DPI)

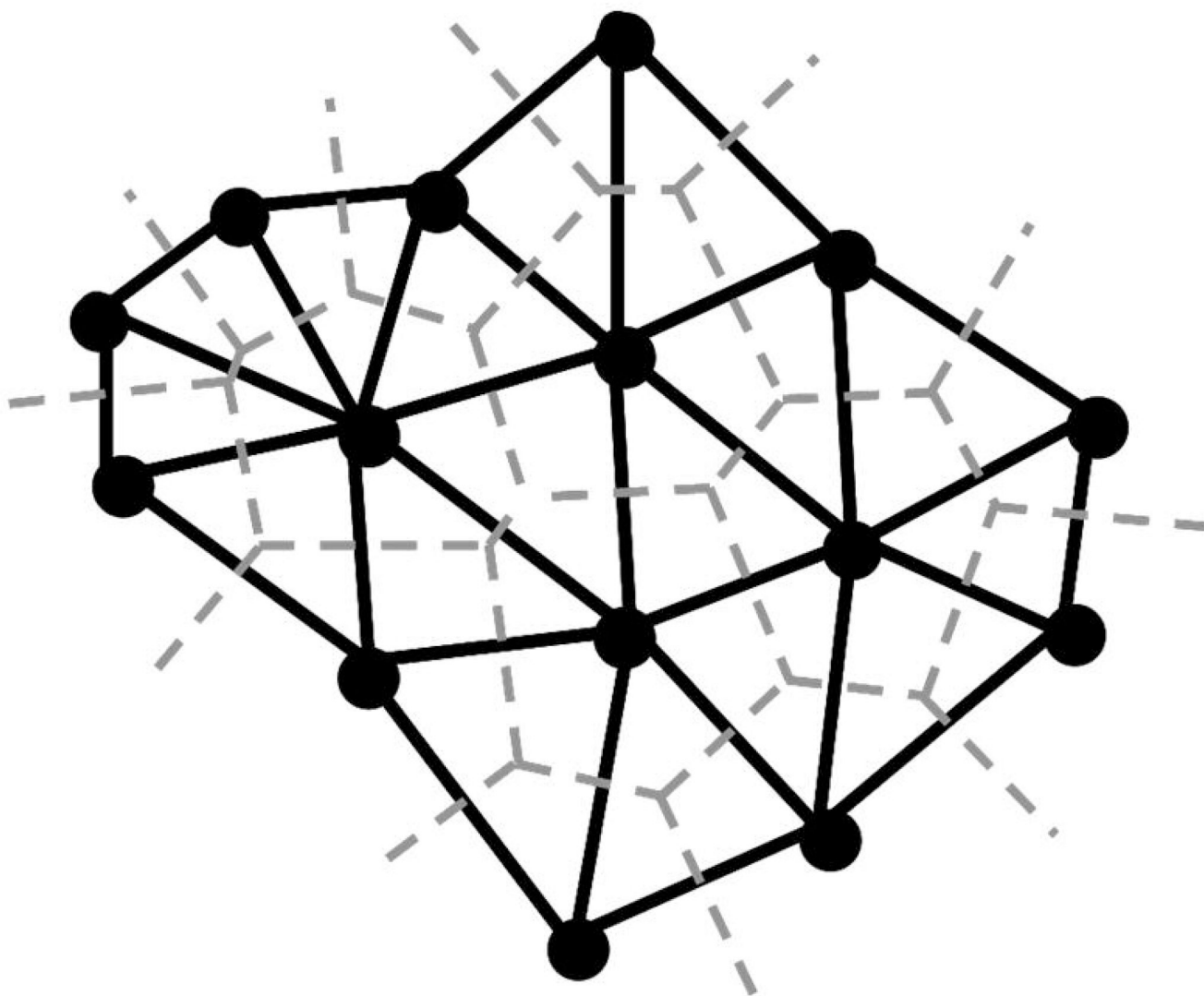


Figure 2. Voronoi/Delaunay tessellation in 2D space (Voronoi polyhedra - dashed line, Delaunay simplices - solid line). 34×29mm (600 × 600 DPI)

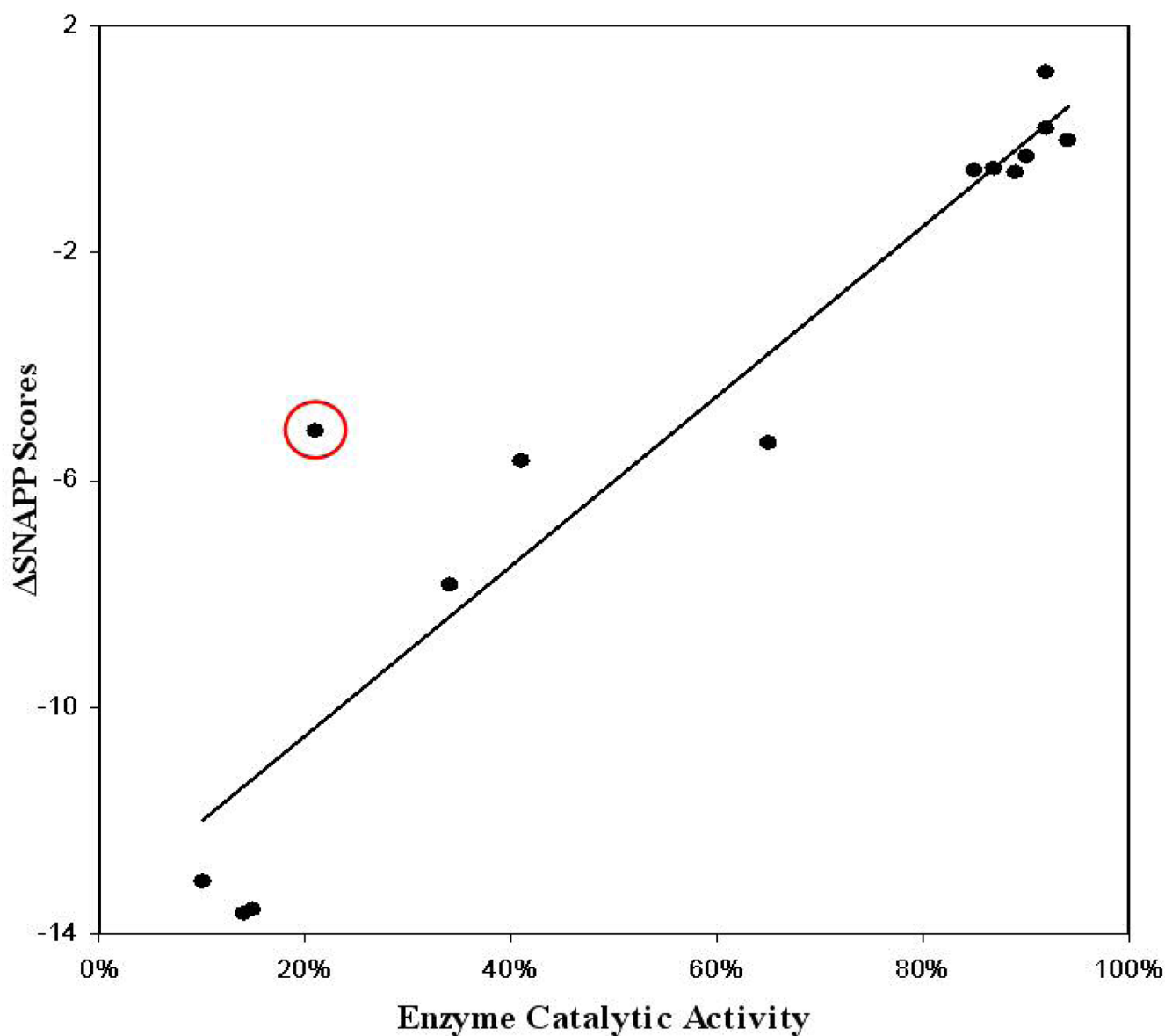


Figure 3. Plot for tethered homodimers reveals a strong correlation between Δ SNAPP score and catalytic activity ($R^2 = 0.88$). After excluding one of the quadruple mutants (in the circle), the R^2 increased to 0.96. 32×29mm (600 × 600 DPI)

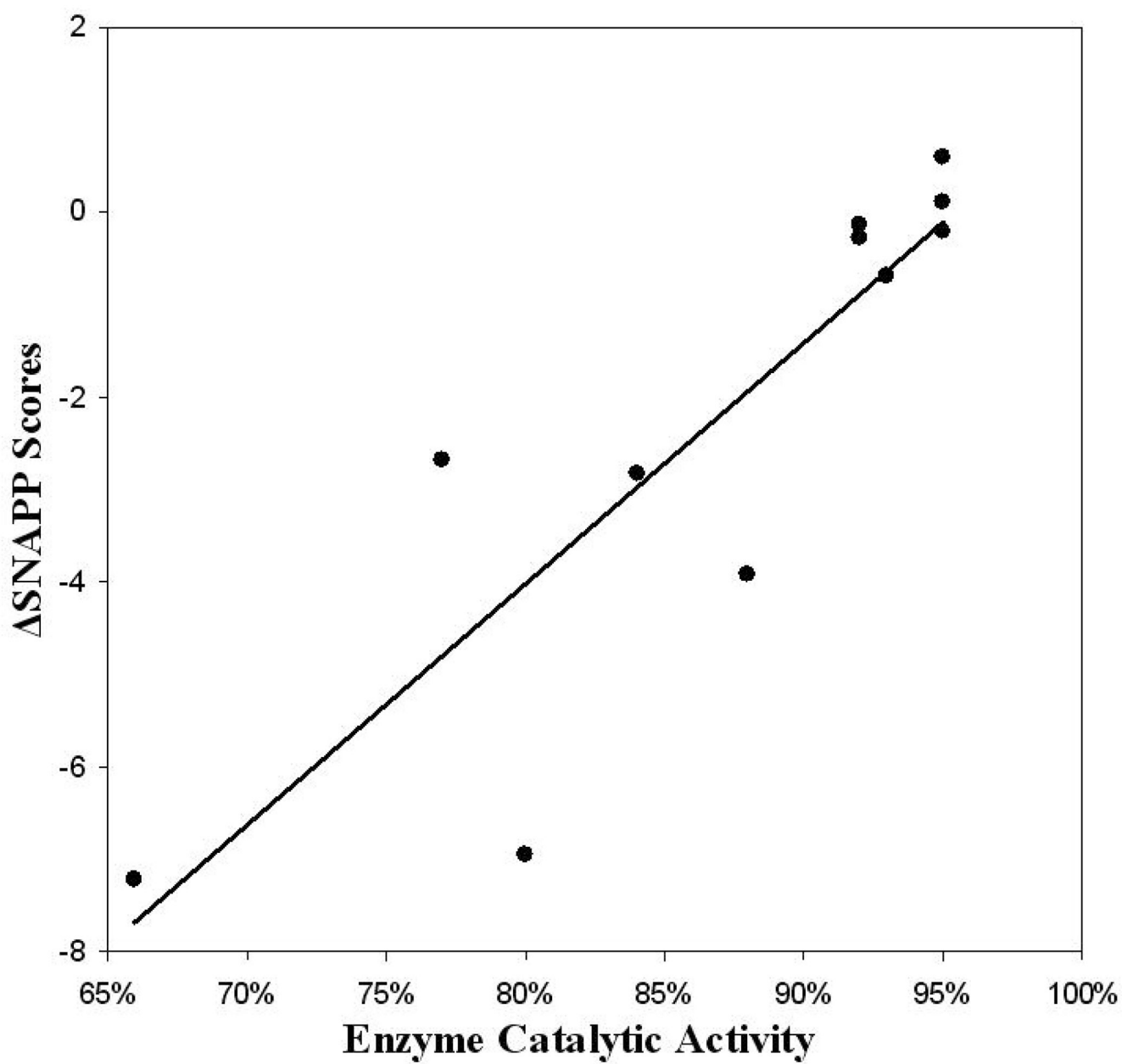


Figure 4.

Plot for tethered heterodimers shows a good correlation between Δ SNAPP scores and enzymatic activity ($R^2 = 0.76$). 31×30mm (600 × 600 DPI)

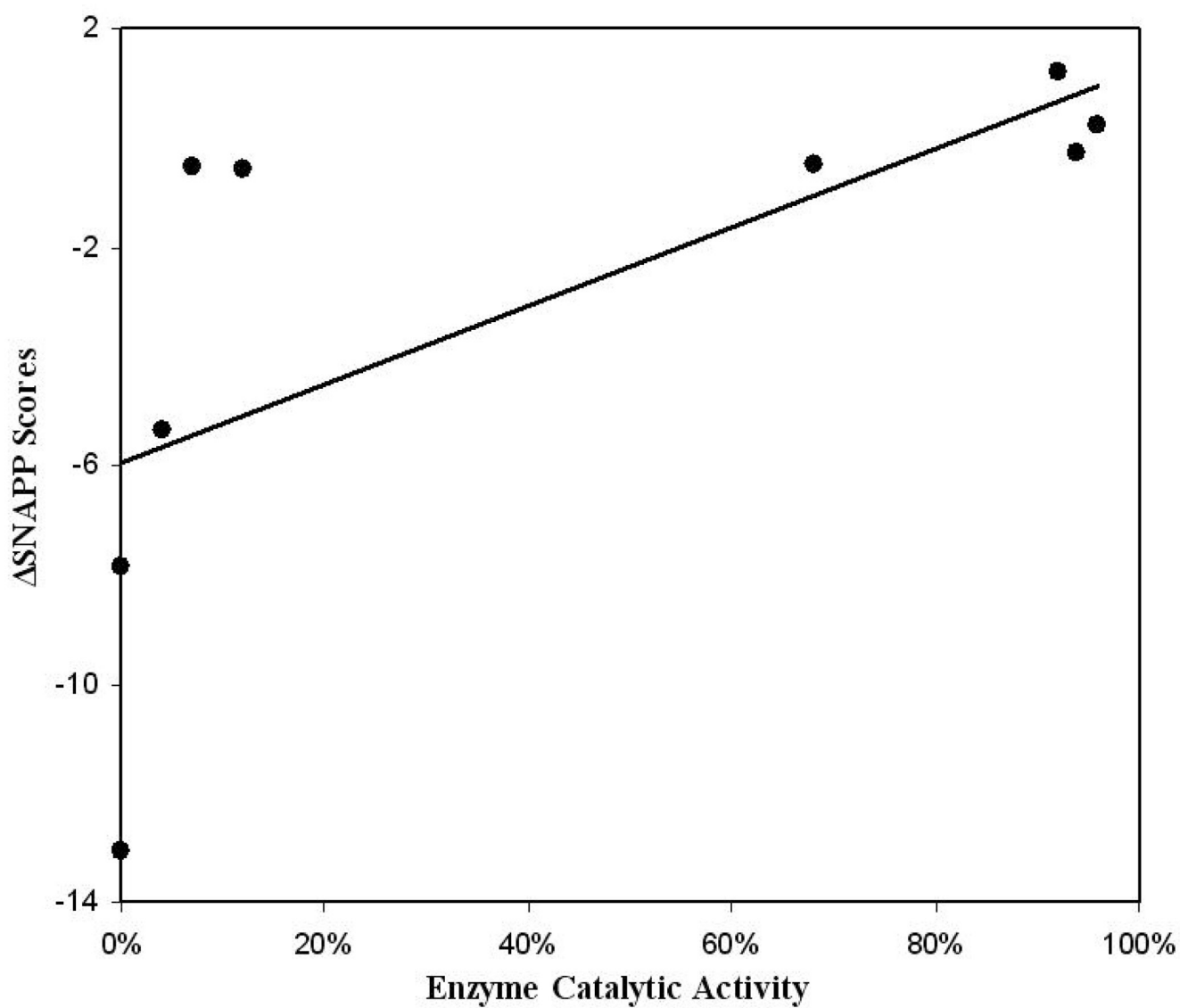


Figure 5.

Plot for non-tethered homodimers indicates that the correlation between Δ SNAPP score and catalytic activity is very weak: R^2 is only 0.45. 32×28mm (600 × 600 DPI)

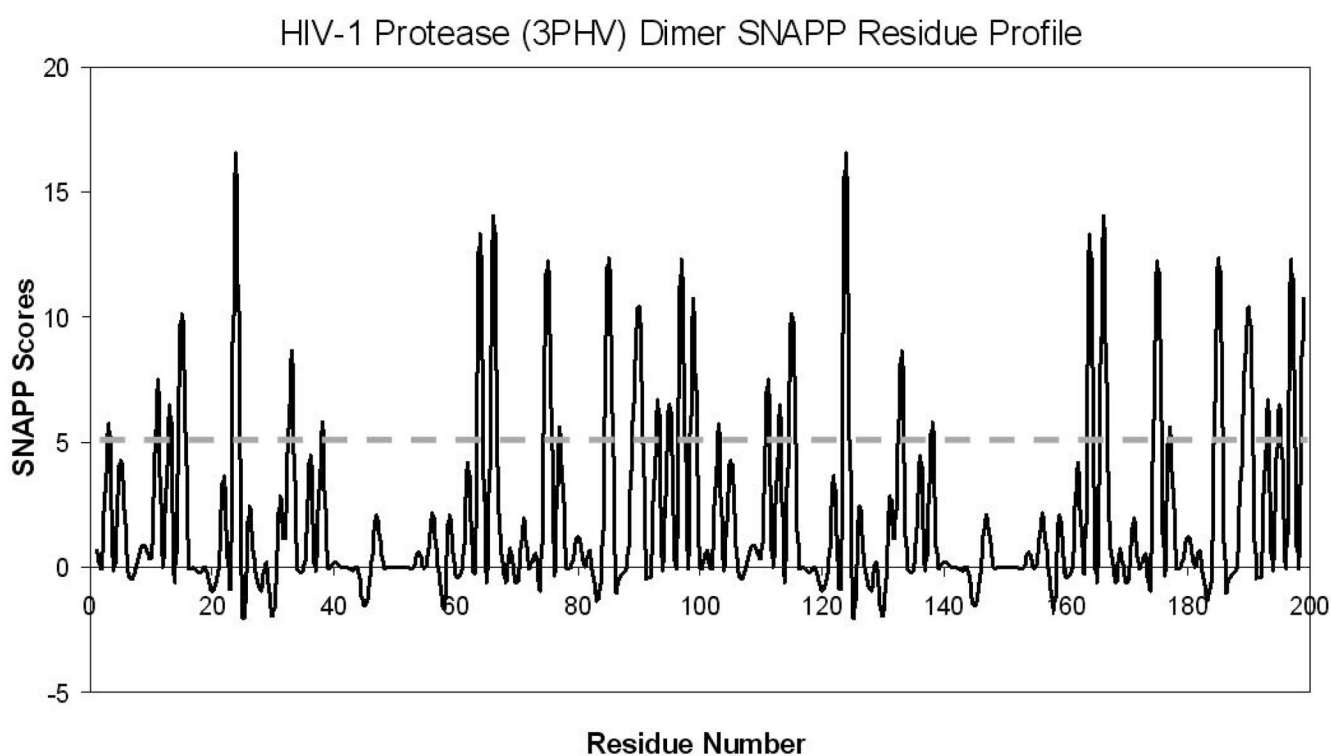


Figure 6.

SNAPP residue profile of HIVP dimer (3PHV). The gray dashed line is a cutoff value: residues with the score higher than the cutoff are buried in the hydrophobic core and stabilize the structure. Our approach indicates that mutations that lower the score may result in unstable (and probably inactive) protease. 44×25mm (600 × 600 DPI)

SNAPP Residue Differential Profile (3PHV Dimer - 3PHV Monomer)

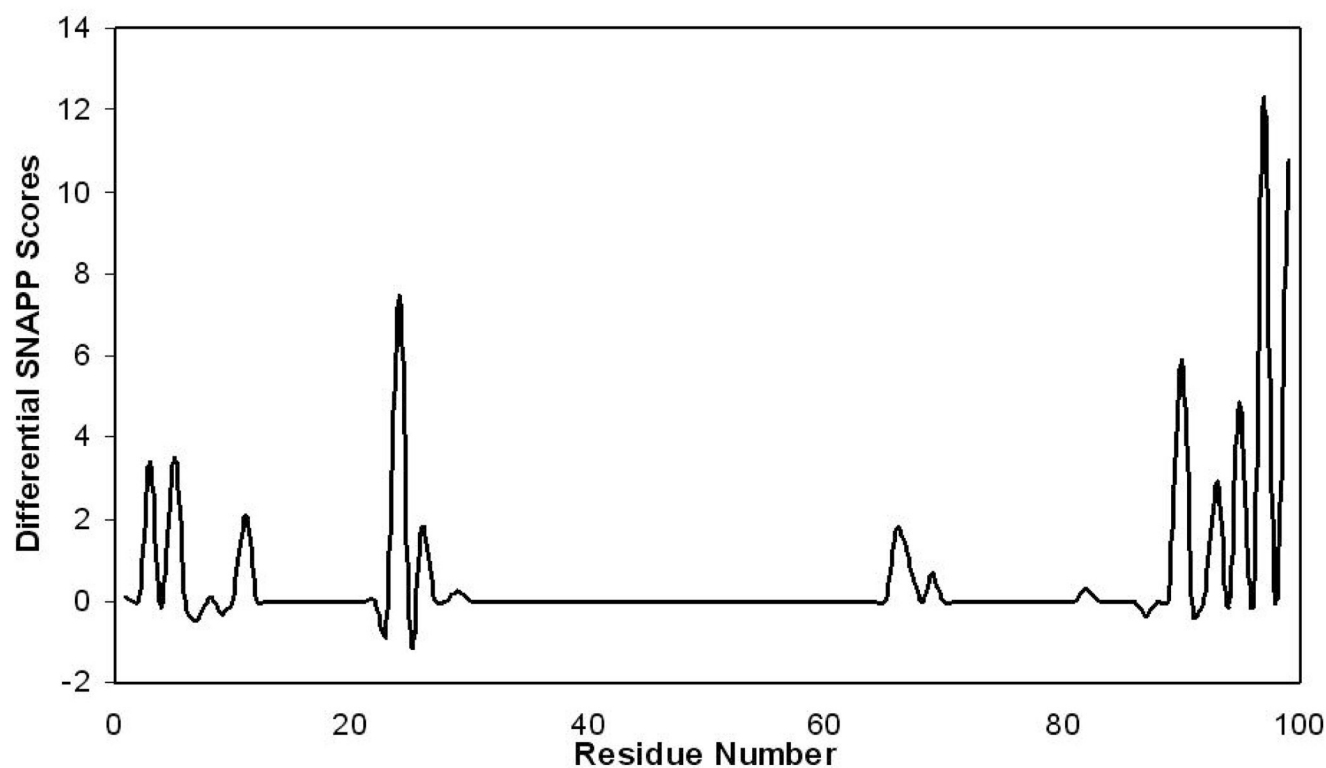


Figure 7. SNAPP residue differential profile for dimeric and monomeric structures was obtained by subtracting monomer residue SNAPP scores from dimer residue SNAPP scores. This plot reveals large changes at the dimer interface (residues 1–4, 96–99) and the active site (around residue 25). 39×27mm (600 × 600 DPI)

SNAPP Residue Differential Profile (1HVC -3PHV)

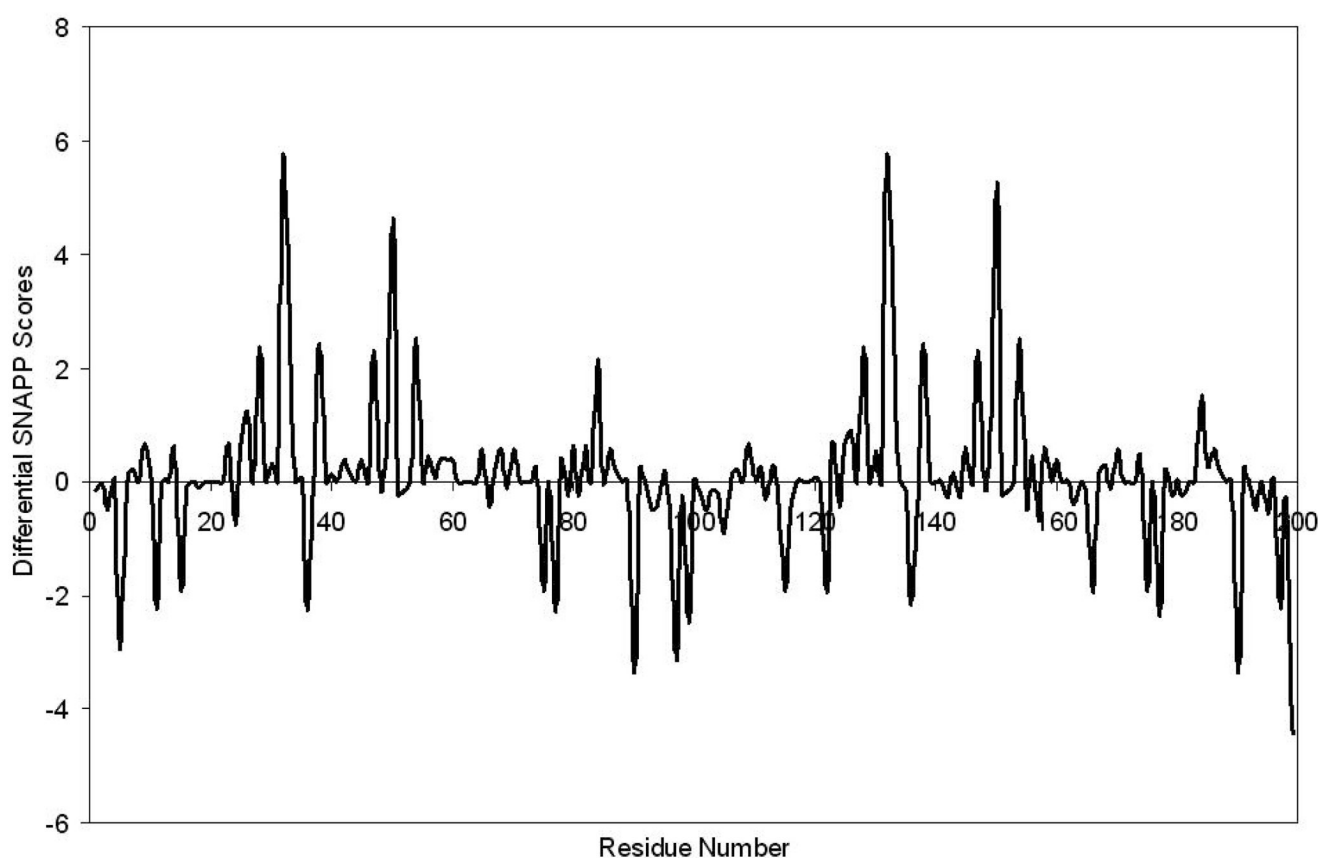


Figure 8.

SNAPP residue differential profile for the liganded and unbound structures was obtained by subtracting residue SNAPP scores of the ligand-free structure from that of the liganded structure. This plot indicates that most of the changes due to ligand binding occur at the dimeric interface (residues 1–4, 96–99), active site (around residue 25) and hydrophobic core (around residue 50). 41×29mm (600 × 600 DPI)

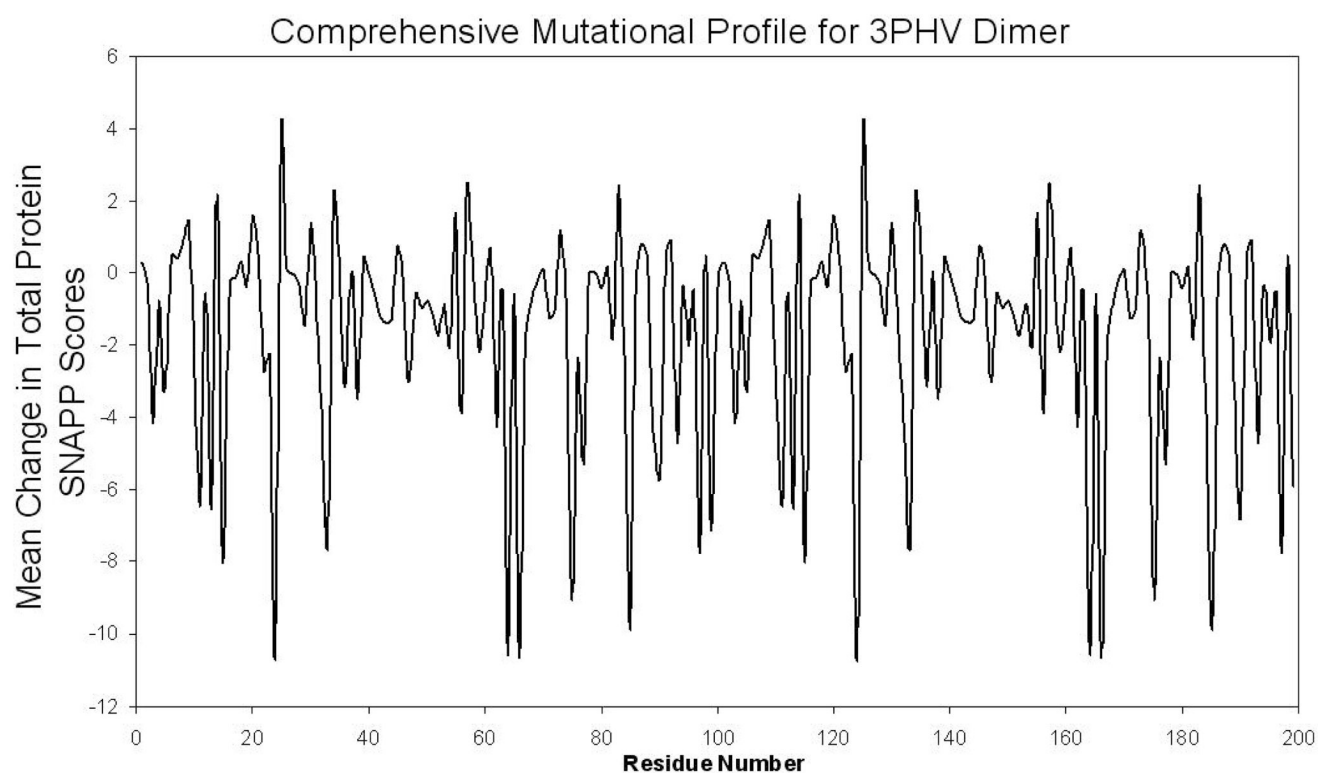


Figure 9. Comprehensive mutational profile plot shows that large fluctuations were observed mostly for residues from the dimeric interface, active site and hydrophobic core. 45×27mm (600 × 600 DPI)

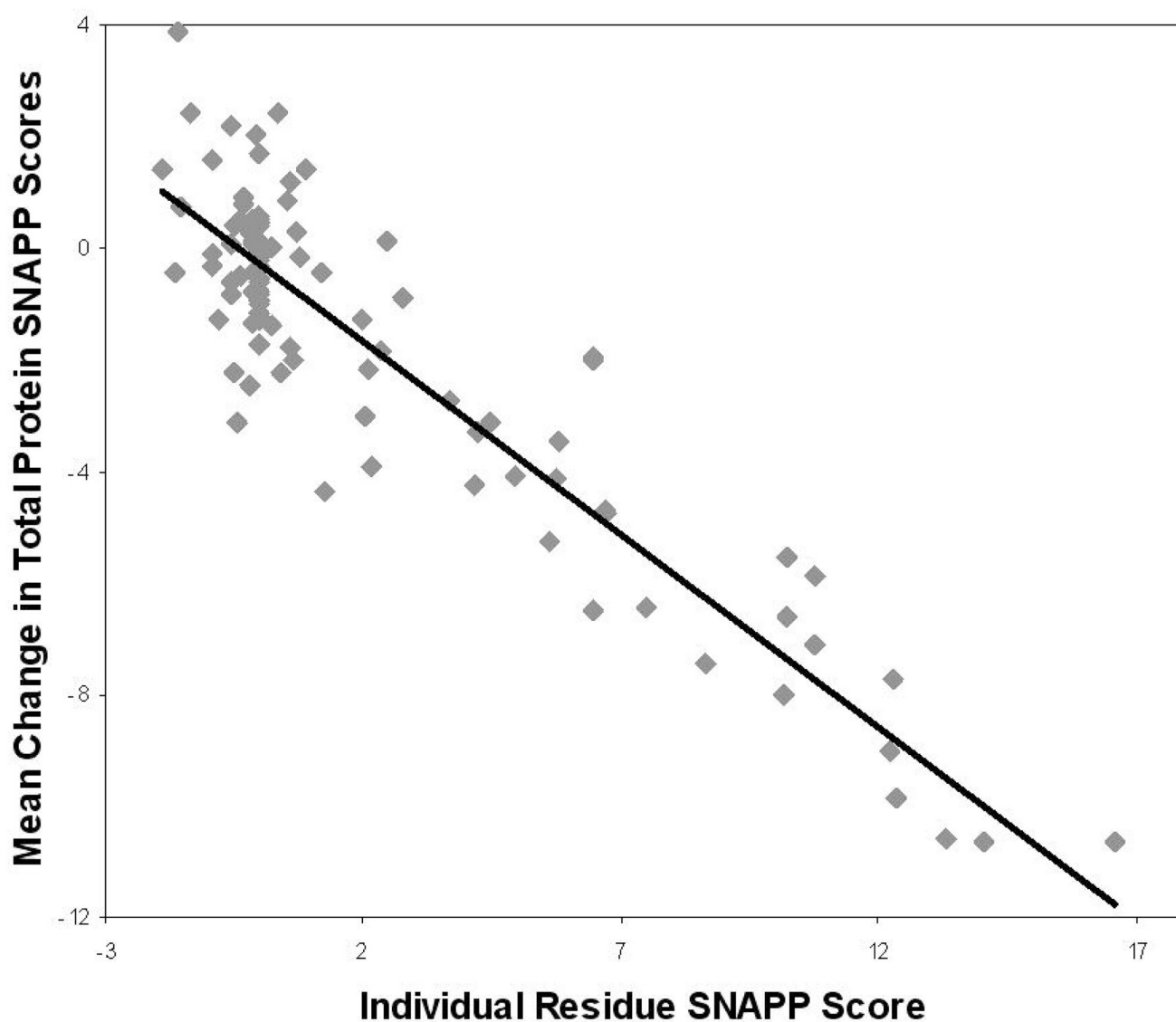


Figure 10.

The plot of the mean change of the total protein (3PHV) SNAPP score resulting from mutations of a residue to the other 19 residue types vs. the wild type residue profile score. The graph reveals a strong inverse relationship, with the correlation coefficient $R^2=0.84$. 38×32mm (600 × 600 DPI)

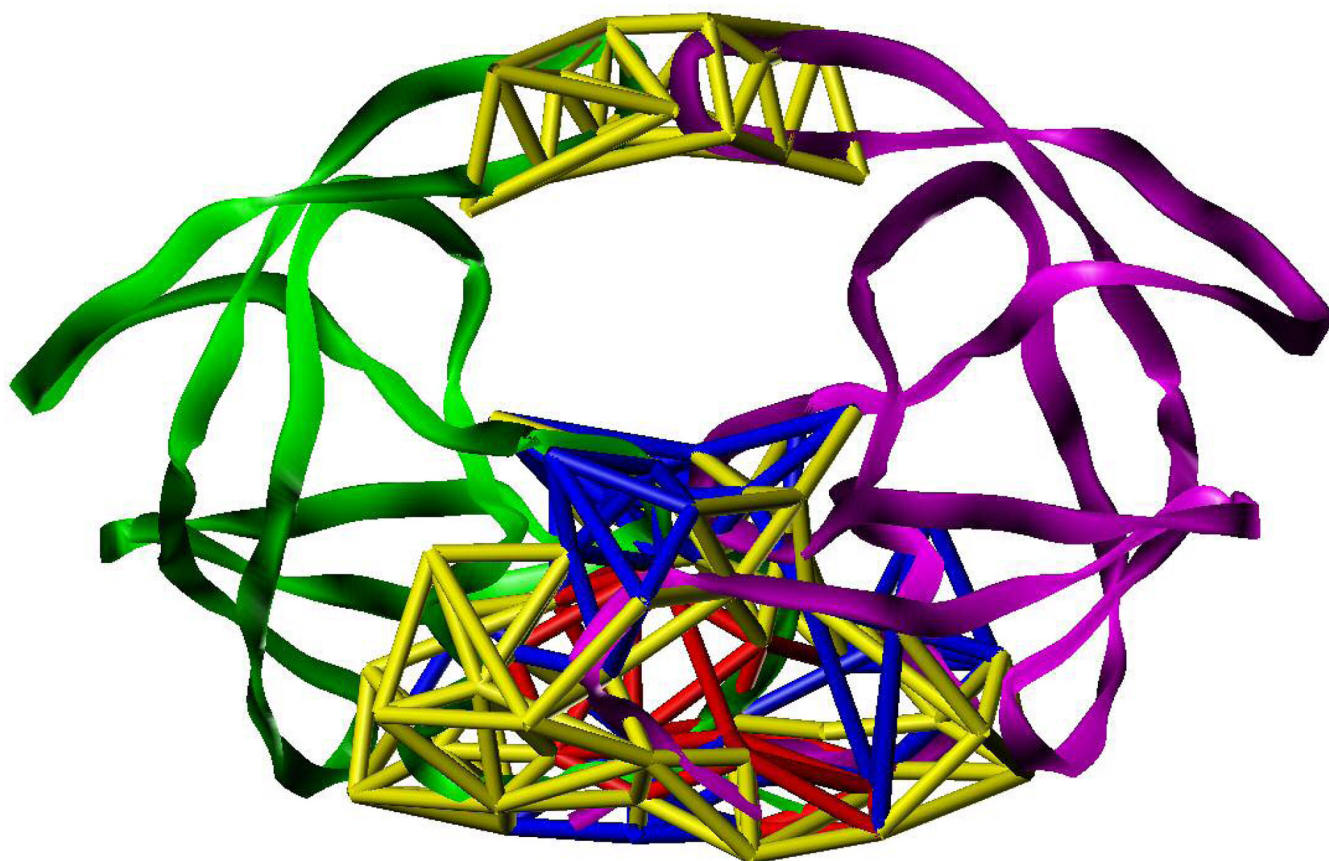


Figure 11.

All of the interfacial tetrahedra (yellow, red, and blue) formed between the two chains (green and magenta) for ligand-free structure (3PHV dimer). There are no tetrahedral contacts between the active site and the flap region while many tetrahedra are formed at the dimeric interface. Red represents high-scoring tetrahedra while the negatively scored tetrahedra are in blue.

45×30mm (600 × 600)

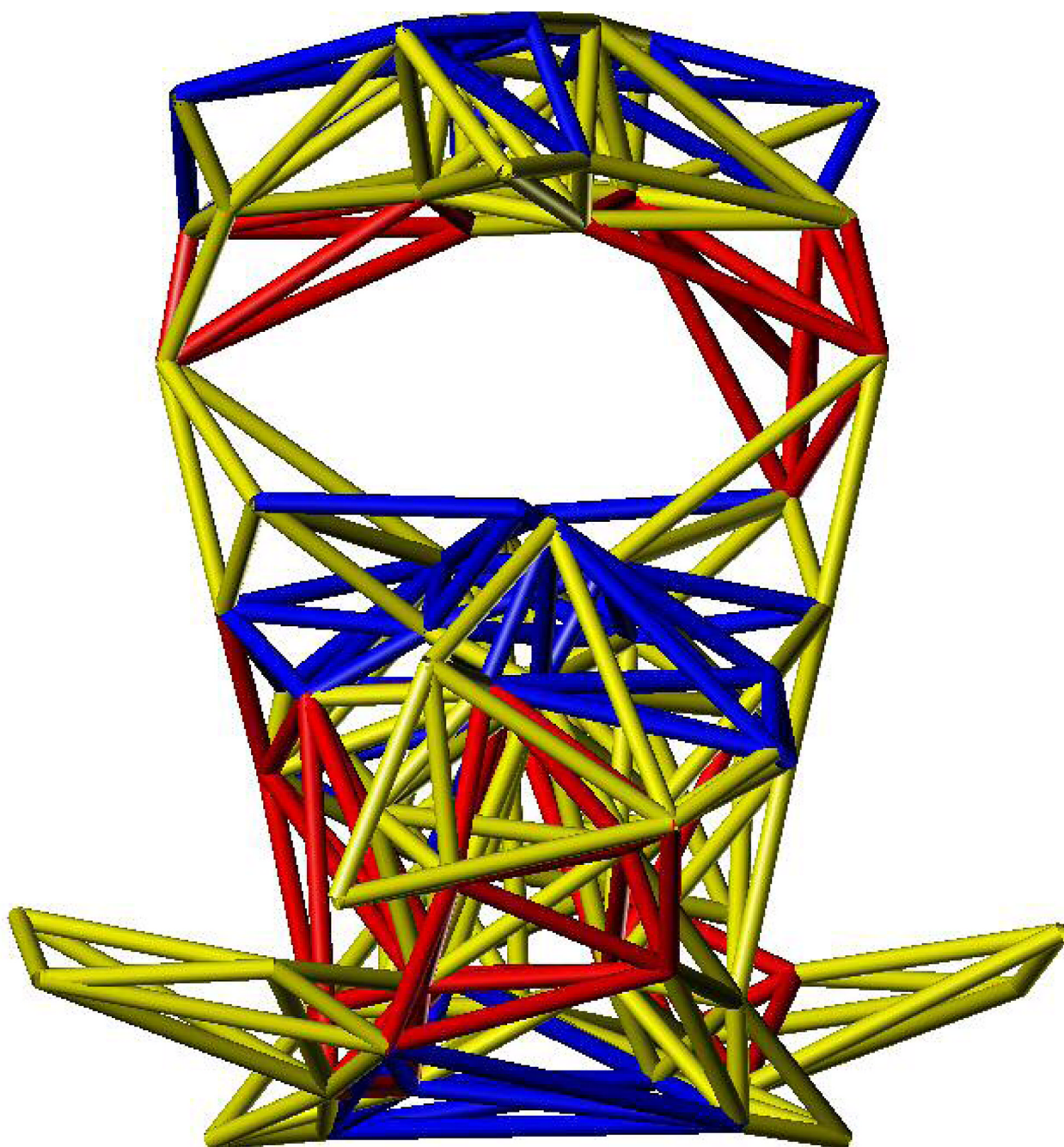


Figure 12.

The interfacial tetrahedra formed between the two subunits of the tethered liganded HIV-1 protease (1HVC). They form an Eiffel Tower-like structure where the hole is the binding pocket and the dimeric interfacial tetrahedra form a very strong “base”. Red represents high-scoring tetrahedra while the negatively scored tetrahedra are in blue. 27×30mm (600 × 600 DPI)

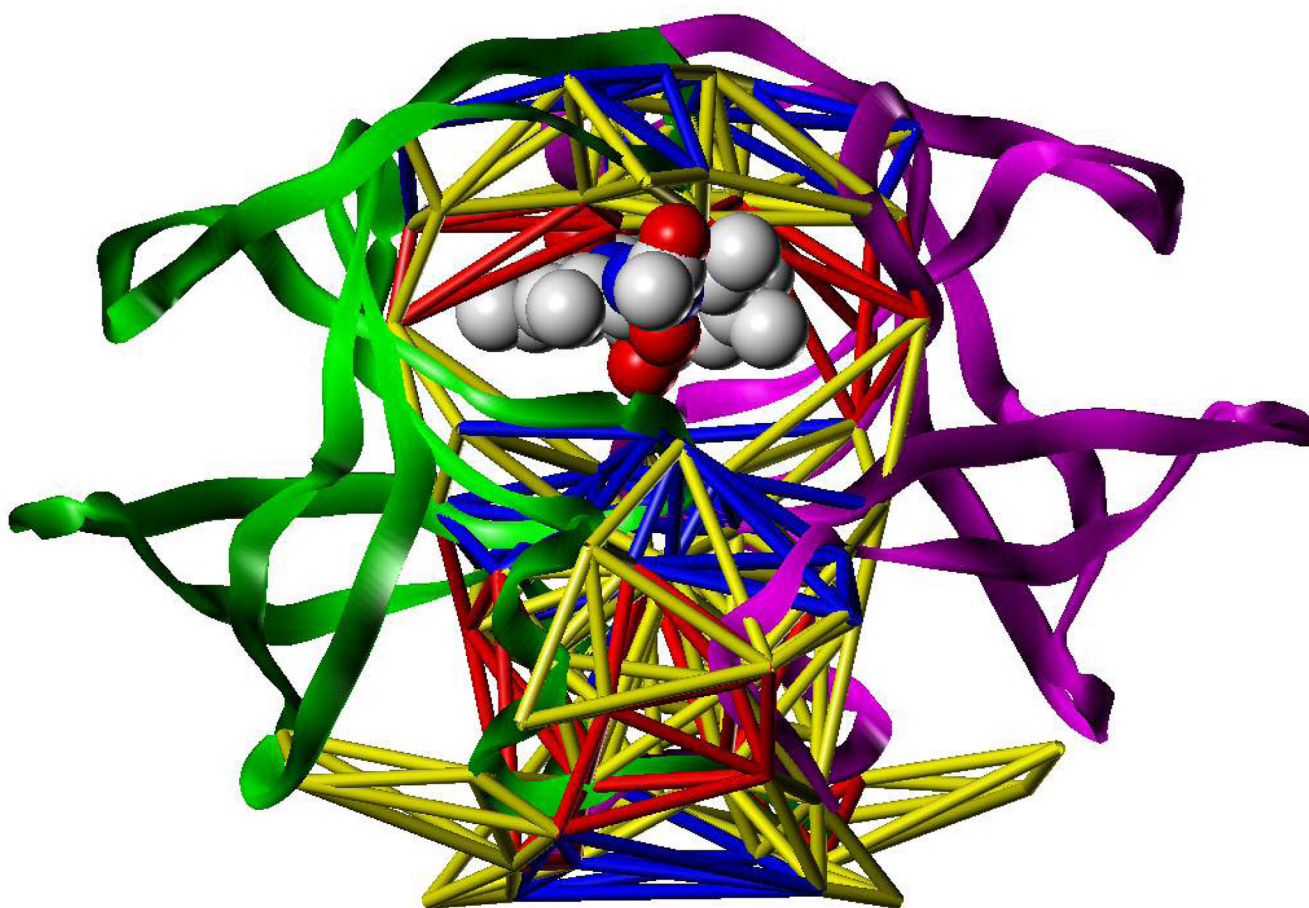


Figure 13.

All of the interfacial tetrahedra (yellow, red and blue) formed between the two chains (green and magenta cartoons) for liganded tethered structure (1HVC). Some tetrahedra are newly formed around the active site and the flap region upon the binding of the ligand (in space filling display). As in the unliganded form, many tetrahedra are formed on the dimeric interface. Red represents high-scoring tetrahedra while the negatively scored tetrahedra are in blue. 43×31mm (600 × 600 DPI)

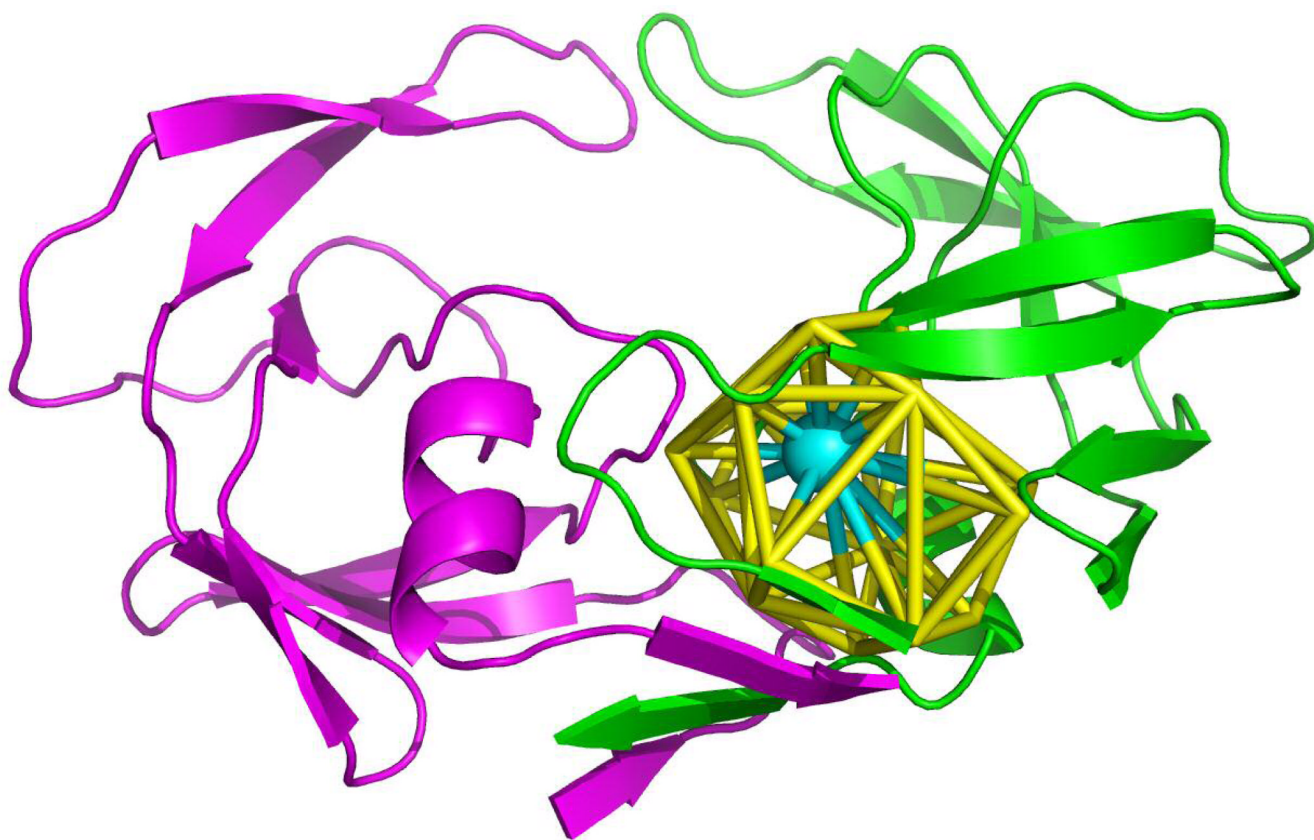


Figure 14. Single residue tessellation of L24 (cyan) and the tetrahedra are yellow. L24 has the highest SNAPP score in the structure and forms 33 tetrahedra with its nearest neighbor residues. The two chains are displayed as cartoons in green and magenta, respectively. 51×33mm (600 × 600 DPI)

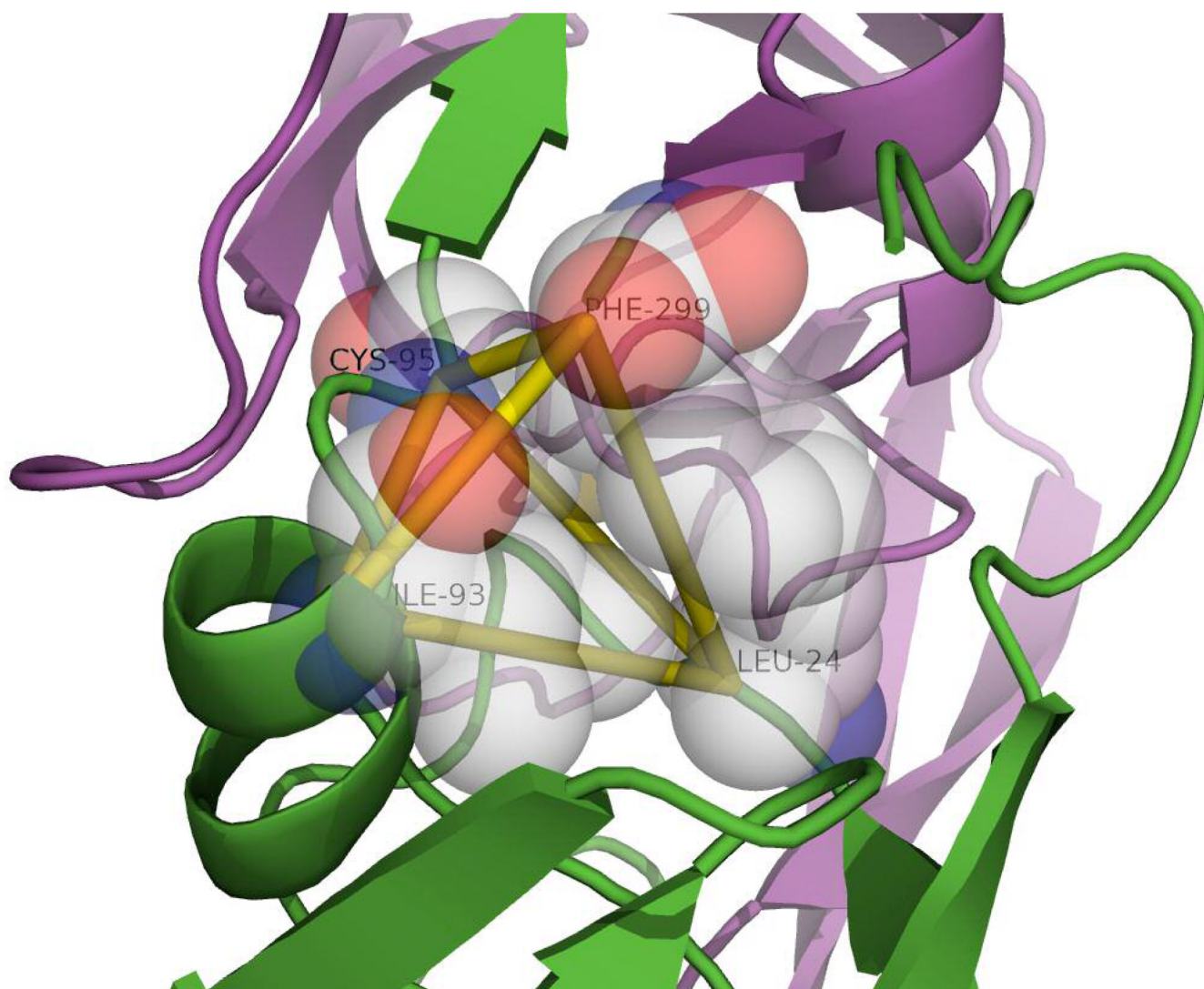


Figure 15.

Four residues (L24-I93-C95-F199 in transparent spheres) form a high-scoring interfacial tetrahedron (yellow sticks). PHE199 is on the dimeric interface and LEU24 is close to the active site. The green and magenta cartoons represent the two chains. 42×33mm (600 × 600 DPI)

Table 1
Experimental Catalytic Activity and SNAPP Score Change for HIV-1 Proteases

Tethered Homodimers		Tethered Heterodimers	
Mutation	% Cleavage	SNAPP Scores	% Cleavage
WT/WT	94%	0	
D25N/D25N	0	1.428	0
1A/1A	92%	1.187	95%
2A/2A	90%	-0.289	92%
3A/3A	65%	-6.362	77%
4A/4A	92%	0.21	95%
96A/96A	85%	-0.555	92%
97A/97A	10%	-13.089	80%
98A/98A	87%	-0.496	95%
98D/98D	89%	-0.567	93%
99A/99A	34%	-7.845	88%
2A3A/2A3A	41%	-5.651	84%
3A4A/3A4A	21%	-5.152	75%
96A/97A/96A/97A	14%	-13.645	66%
97A/98A/97A/98A	15%	-13.585	84%
Non-tethered Dimers			
96A	7%	-0.555	1A
97A	0	-13.089	2A
98A	68%	-0.496	3A
98D	12%	-0.567	4A
99A	0	-7.845	