Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 305 (2003) 322–326
BBRC
www.elsevier.com/locate/ybbrc

Comprehensive mutagenesis of HIV-1 protease: a computational geometry approach☆

Majid Masso and Iosif I. Vaisman*

Bioinformatics and Computational Biology, School of Computational Sciences, George Mason University, 10900 University Blvd., MSN 5B3, Manassas, VA 20110, USA

Received 28 March 2003

Abstract

A computational geometry technique based on Delaunay tessellation of protein structure, represented by C_α atoms, is used to study effects of single residue mutations on sequence–structure compatibility in HIV-1 protease. Profiles of residue scores derived from the four-body statistical potential are constructed for all 1881 mutants of the HIV-1 protease monomer and compared with the profile of the wild-type protein. The profiles for an isolated monomer of HIV-1 protease and the identical monomer in a dimeric state with an inhibitor are analyzed to elucidate changes to structural stability. Protease residues shown to undergo the greatest impact are those forming the dimer interface and flap region, as well as those known to be involved in inhibitor binding.

Keywords: Statistical geometry; Delaunay tessellation; Molecular modeling; Protein structure; Computational mutagenesis; HIV-1 protease

A computational geometry technique based on Delaunay tessellation of protein structure, represented by α-carbons in 3D space, provides an objective and robust definition of four nearest-neighbor amino acid residues as well as a four-body statistical potential function [1–3]. Utilizing this function, a residue score characterizing sequence–structure compatibility can be calculated for the purposes of obtaining a total protein potential and constructing a 3D–1D potential profile. By replacing individual residues in a protein, recalculating the profiles, and comparing the new and original profiles, we can obtain valuable information for the computational study of effects of mutations on protein structure and function. In this paper, we discuss an application of this approach to a comprehensive mutational analysis of HIV-1 protease.

HIV-1 protease is responsible for effective assembly of budding virions by cleavage of the gag and gag-pol polyprotein precursors that encode vital structural and replicative proteins. Inhibitors of HIV-1 protease, in combination with other anti-HIV compounds, are highly effective in reducing levels of viral replication. However, due to the high error rate of HIV-1 reverse transcriptase, there exists the potential for development of residue mutations leading to inhibitor resistance. The data collected from our computational mutagenesis allow for a systematic approach to predicting conformational changes in HIV-1 protease resulting from such mutations.

Functional HIV-1 protease exists as a dimer of identical 99 amino acid subunits with an intermolecular twofold axis of symmetry through the substrate-binding pocket. Additionally, each monomer contains an approximate intramolecular twofold axis of symmetry [4]. A four-stranded β-sheet consisting of the interdigitated N- and C-termini (residues P1-T4 and C95-F99, respectively) of each subunit forms the dimer interface [4,5]. The active site consists of the triad D25-T26-G27 and both D25 residues of the active dimer together mediate polyprotein catalysis [4]. A hydrophobic extended turn, or "flap," is formed by residues M46-V56 [6]. Although the two flaps cover the active site of the unliganded dimer, they are capable of distorting to accommodate and interact with substrate molecules by virtue of their glycine rich composition [4,6].

☆ *Abbreviations:* WT, wild-type; CMP, comprehensive mutational profile; PDB, Protein Data Bank.

* Corresponding author. Fax: 1-703-993-8401.
E-mail address: ivaisman@gmu.edu (I.I. Vaisman).

0006-291X/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0006-291X(03)00760-5

Within our modeling framework, the residue present at each position of the wild-type (WT) HIV-1 protease monomer is permuted with each of the other 19 naturally occurring amino acids. The procedure yields a rich source of statistical potential data for each of the 1881 mutant protein chains, similar to that obtained for the WT monomer. A comprehensive mutational profile (CMP) for the HIV-1 protease monomer is developed by: (1) taking the difference between mutant and WT total potentials for each of the 1881 mutants; (2) segregating the mutational data according to residue position; and (3) averaging the resulting 19 data values at each of the 99 positions. Clustering of the 19 data values at each position, based on whether the corresponding residue replacements represent conservative or non-conservative mutations of the WT residue, is carried out for an additional level of detail. Similarly, clustering according to a 3-letter alphabet, by segregating hydrophobic, charged, and polar residue substitutions and averaging the values within each of these subgroups, is also provided. Returning to WT HIV-1 protease, 3D–1D potential profiles are obtained for a dimer with a bound inhibitor as well as for an isolated monomer from the same structure. A comparison of these profiles reveals the impact that dimerization and inhibitor binding have on structural stability according to the model.

Materials and methods

Tessellation. In this construction, a protein structure is represented by a set of points in 3D space, whose coordinates are those of the C_α atoms of the residues forming the protein [1–3]. Voronoi tessellation partitions the 3D space into a set of convex polyhedra, each containing a single C_α atom, such that the interior points of each polyhedron are closer to their corresponding C_α atom than any other in the system. Delaunay tessellation is obtained by connecting sets of four C_α atoms whose Voronoi polyhedra meet at a common vertex. The result is an aggregate of space-filling irregular tetrahedra, or Delaunay simplices. Vertices associated with each Delaunay simplex objectively define a set of four nearest-neighbor residues without any explicit dependence on an adjustable distance parameter. The 2D representation of Voronoi and Delaunay tessellations is shown in Fig. 1. Given a Protein Data Bank (PDB) [7] coordinate file, Delaunay tessellation of the protein structure is performed using the Quickhull algorithm [8].

Statistical potential. Assuming that residue composition of Delaunay simplices is order independent, the theoretical maximum number of all possible combinations of quadruplets of 20 amino acids forming such simplices is 8855 [1–3]. Statistical analysis of the unbiased preference for four residues to be clustered spatially in a folded protein proceeds as follows. The log-likelihood for each quadruplet is defined as $q_{ijkl} = \log(f_{ijkl}/p_{ijkl})$, where f_{ijkl} represents the frequency of quadruplets containing residues i, j, k, l in a representative training set of high-resolution protein structures with low primary sequence identity, and p_{ijkl} is the frequency of random occurrence of the quadruplet [1–3]. Total potential of a protein is calculated by summing the log-likelihoods of all Delaunay simplices formed by the tessellation [1,2]. By summing the log-likelihoods of all simplices in which a particular residue participates, individual residues are each assigned a score, yielding a 3D–1D potential profile [1,2].

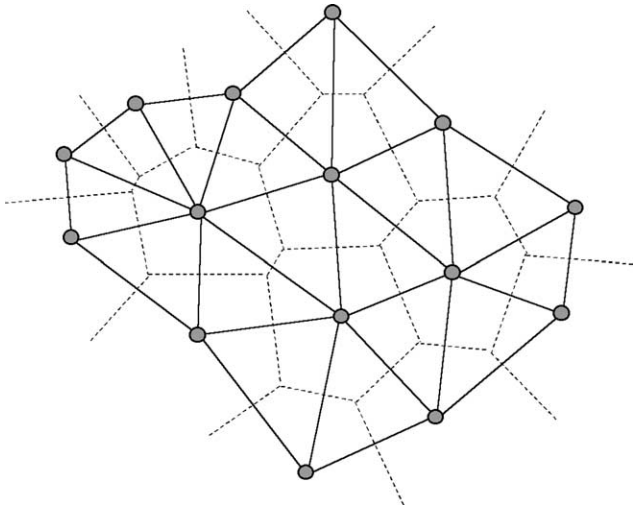

Fig. 1. Dual tessellations in 2D space: the dashed/solid lines indicate Voronoi and Delaunay tessellations, respectively.

Results and discussion

HIV-1 protease potential profile

The total statistical potential of a single subunit of WT HIV-1 protease (PDB ID: 3phv [9]) is calculated to be 27.93 and the potential profile of the monomer is shown in Fig. 2. A few features of this 3D–1D plot are worth noting. First, the shape of the graph reveals the approximate intramolecular twofold axis of symmetry of the monomer. Next, the active site D25 residue as well as residues forming the flap region have low statistical potential values, by virtue of the fact that stability is

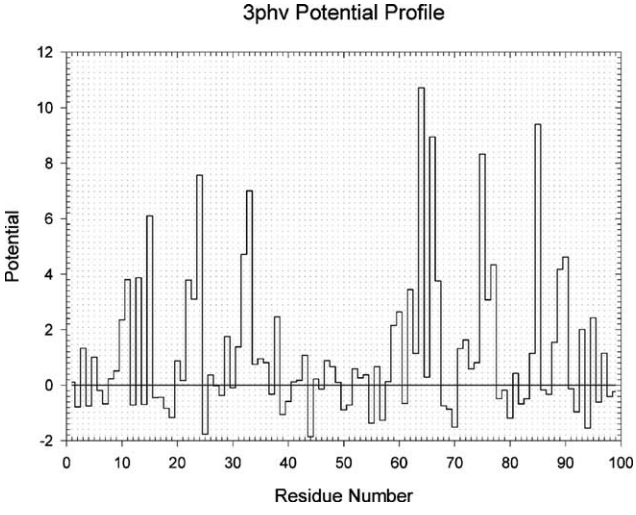

Fig. 2. 3D–1D potential profile of a wild-type HIV-1 protease monomer (PDB ID: 3phv). Highest scoring residues tend to be in the hydrophobic core, while lower scoring residues are exposed. A local minimum occurs at the active site D25. The total potential of the protease is 27.93.

affected by the lack of interaction with either an accompanying flap (as part of a dimeric complex) or a substrate molecule. Finally, residues with the highest potentials are mostly hydrophobic and serve to maintain stability within the core of the protein chain.

HIV-1 protease comprehensive mutational profile

The initial step toward developing the CMP involves mutating 19 times the residue present at each position of the 99 amino acid long HIV-1 protease monomer. As a result, a 20×99 matrix of total statistical potentials for all possible single residue mutants is produced. Each of the 99 columns is labeled with the corresponding residue present in the primary sequence of WT HIV-1 protease, and each of the 20 rows is labeled with an amino acid chosen to replace the given residue in the primary sequence. Subtracting the WT potential from each cell yields the difference between mutant and WT total potentials. Finally, by averaging the values in each column, the CMP is obtained (Fig. 3).

Comparison of the HIV-1 protease potential profile in Fig. 2 and the corresponding CMP in Fig. 3 reveals a strong inverse correlation, as depicted in Fig. 4. Analysis of the plots shows that each residue potential in the 3D–1D profile is dampened by the mean of the 19 differences between mutant and WT total potentials calculated for all substitutions of the given residue. To determine whether this phenomenon is a property unique to native-fold proteins, we reshuffled the residues in the primary sequence of 3phv and repeated the analysis. The correlation coefficient for this random sequence is significantly lower than that for the native one ($R^2 = 0.74$).

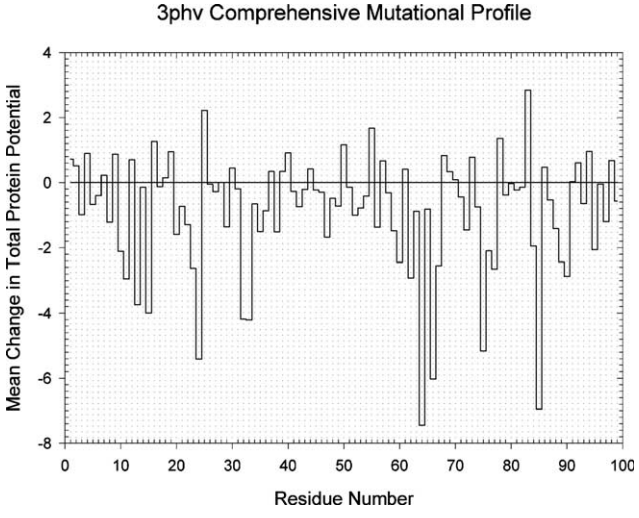

Fig. 3. Comprehensive mutational profile of an HIV-1 protease monomer. For each residue, the graph reflects the mean difference in total potential from wild-type protease resulting from all possible substitutions at the given position.

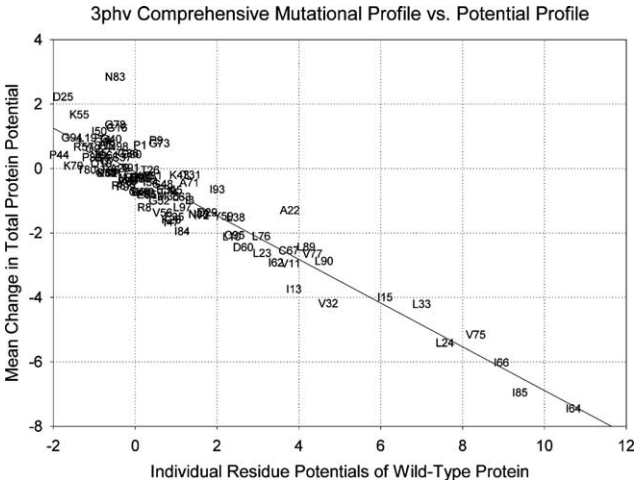

Fig. 4. Correlation of the CMP with the individual residue potentials of wild-type HIV-1 protease. The graph reveals a strong inverse relationship, with correlation coefficient $R^2 = 0.88$.

Clustering schemes applied to the comprehensive mutational profile

A similar analysis is conducted by segregating the 20 amino acids into the following 6 groups, as described by Dayhoff et al. [10]: (A, S, T, G, and P), (V, L, I, and M), (R, K, and H), (D, E, N, and Q), (F, Y, and W), and (C). Substitution of a residue with another one from within the same group represents a conservative mutation; otherwise, the mutation is non-conservative. A CMP based on this type of clustering is shown in Fig. 5A. Examples of residues that warrant further investigation include A22 and A71, given the much greater negative impact that conservative mutations have over non-conservative mutations. By comparing Figs. 2 and 5A, it is clear that a strong inverse correlation exists between the 3D–1D potential profile of WT HIV-1 protease and the CMP for non-conservative substitutions of the residues ($R^2 = 0.88$, not shown). No correlation is evident with respect to conservative mutations ($R^2 = 0.16$, not shown).

Another approach involves clustering the amino acid substitutions according to a 3-letter alphabet based on side chain polarity: F (hydrophobic residues A, V, L, I, M, P, and F); L (charged residues D, E, R, and K); and P (polar residues N, Q, W, S, T, G, C, H, and Y). The CMP for this scenario (Fig. 5B) is obtained by averaging the values within each subgroup. It is interesting to note that most hydrophobic residue positions within the HIV-1 protease core exhibit steep decreases in potential, regardless of the type of residue substitution.

Changes in structural stability due to dimerization and/or ligand binding

Comparing the 3D–1D profile of an isolated HIV-1 protease monomer with that of the same monomer in a

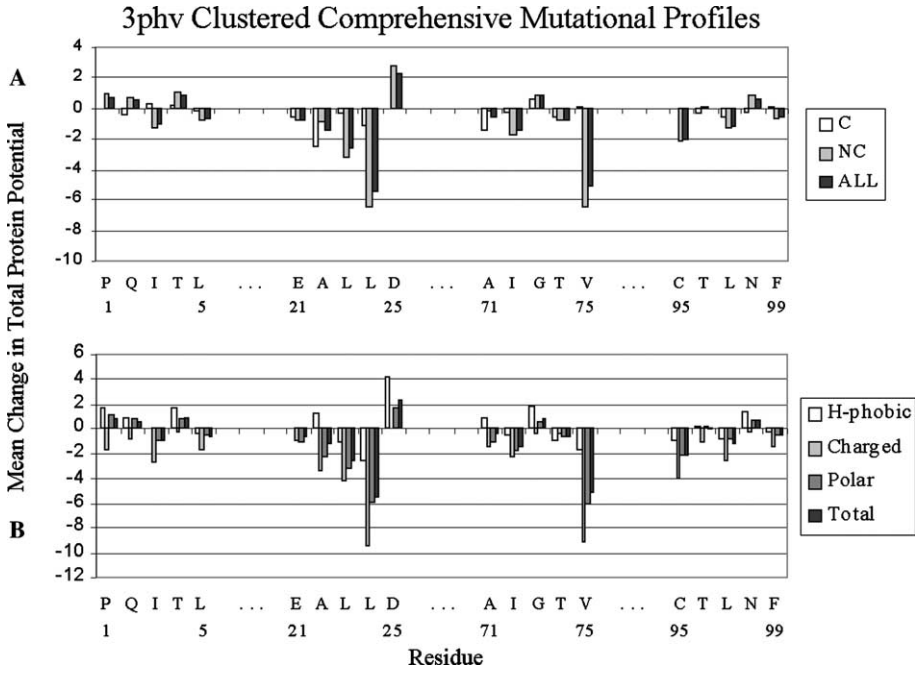

Fig. 5. Clustering the CMP data, based on conservative versus non-conservative substitutions of the wild-type residue at each position (A) and a 3-letter alphabet (F, hydrophobic; L, charged; and P, polar) characterizing the polarity of the side chains of the substitutions (B).

homodimeric configuration with a bound inhibitor provides valuable information for the study of protein–protein and ligand–receptor interfaces. First, the same procedure used to obtain the potential profile for the 3phv monomer is applied to a dimer–inhibitor complex (PDB ID: 1g35 [11]). A 3D–1D plot is created for the 1g35A chain following tessellation of the complex. Next, a potential profile is created for the 1g35A chain, tessellated as an isolated monomer.

The difference between these potential profiles (Fig. 6) reveals the change in stability that affects residues involved in the dimerization of the protein chain as well as the binding of an inhibitor. For example, the overwhelming majority of residues forming both the dimer interface and the flap region exhibit increases in potential following dimerization, including Q2, T4, I47-I54, T96, L97, and F99. With the exception of Q2, these residues are also characteristically all hydrophobic. Similar increases in stability due to inhibitor binding are evident for the active site residues D25, T26, and G27, as well as for the surrounding hydrophobic residues L24 and A28.

The computational geometry approach to protein structure analysis is used to perform a comprehensive mutagenesis of HIV-1 protease. A strong inverse correlation between the 3D–1D potential profile of WT HIV-1 protease and the CMP was found. Clustering the comprehensive mutational profile based on the type of substitution (conservative vs. non-conservative) and on a reduced 3-letter alphabet (hydrophobic, polar, and charged) provides further insight into evolutionary influences on protein structure. While a strong inverse correlation exists between the 3D–1D potential profile of WT HIV-1 protease and the CMP for non-conservative substitutions, no correlation is evident with the conservative substitutions. Finally, comparing the 3D–1D potential profile of an isolated monomer of WT HIV-1 protease with that of the monomer as part of a homodimer with a bound inhibitor reveals changes in

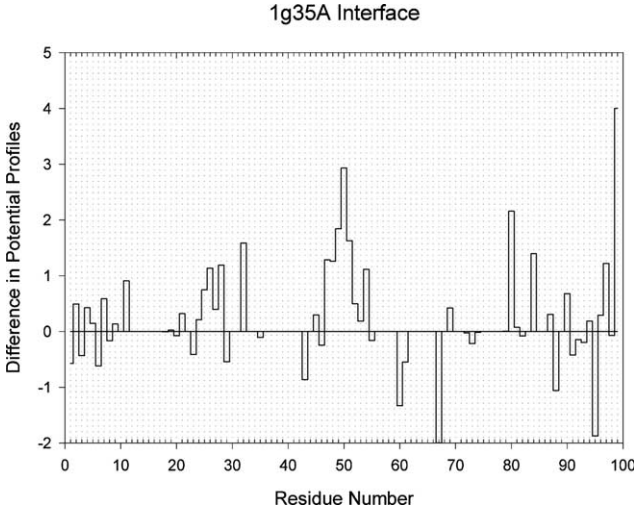

Fig. 6. Difference in potential profiles between an HIV-1 protease monomer in a dimeric configuration with an inhibitor and an isolated monomer from the same structure (PDB ID: 1g35). Residues located at the dimer interface, along the flap region, and within the active site of the monomer experience an increase in potential following dimerization and ligand binding, reflecting altered stability.

structural stability defined by the interface between the monomers that form the dimer.

Acknowledgments

This work was supported in part by a pilot grant from the UNC Center for AIDS Research. We thank Zhibin Lu for his help with writing codes used in this study.

References

[1] R.K. Singh, A. Tropsha, I.I. Vaisman, Delaunay tessellation of proteins: four body nearest neighbor propensities of amino acid residues, J. Comput. Biol. 3 (1996) 213–222.
[2] I.I. Vaisman, A. Tropsha, W. Zheng, Compositional preferences in quadruplets of nearest neighbor residues in protein structures: statistical geometry analysis, Proc. IEEE Symp. Intell. Syst. (1998) 163–168.
[3] W. Zheng, S.J. Cho, I.I. Vaisman, A. Tropsha, A new approach to protein fold recognition based on Delaunay tessellation of protein structure, in: R.B. Altman, et al. (Eds.), Biocomputing '97, World Scientific, Singapore, 1997, pp. 486–497.
[4] A. Wlodawer, J. Vondrasek, Inhibitors of HIV-1 protease: a major success of structure assisted drug design, Annu. Rev. Biomol. Struct. 27 (1998) 249–284.
[5] M.D. Shultz, J. Chmielewski, Probing the role of interfacial residues in a dimerization inhibitor of HIV-1 protease, Bioorg. Med. Chem. Lett. 9 (1999) 2431–2436.
[6] W. Shao, L. Everitt, M. Manchester, D.D. Loeb, C.A. Hutchison III, R. Swanstrom, Sequence requirements of the HIV-1 protease flap region determined by saturation mutagenesis and kinetic analysis of flap mutants, Proc. Natl. Acad. Sci. USA 94 (1997) 2243–2248.
[7] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242, URL: http://www.pdb.org/.
[8] C.B. Barber, D.P. Dobkin, H.T. Huhdanpaa, The quickhull algorithm for convex hulls, ACM Trans. Math. Software 22 (1996) 469–483, URL: http://www.geom.umn.edu/software/qhull/.
[9] R. Lapatto, T. Blundell, A. Hemmings, J. Overington, A. Wilderspin, S. Wood, J.R. Merson, P.J. Whittle, D.E. Danley, K.F. Geoghegan, et al., X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes, Nature 342 (1989) 299–302.
[10] M.O. Dayhoff, R.M. Schwartz, B.C. Orcut, in: M.O. Dayhoff (Ed.), Atlas of Protein Sequence and Structure, 5, Nat. Biomedical Research Fdn, Washington, DC, 1978, pp. 345–352.
[11] W. Schaal, A. Karlsson, G. Ahlsen, J. Lindberg, H.O. Andersson, U.H. Danielson, B. Classon, T. Unge, B. Samuelsson, J. Hulten, A. Hallberg, A. Karlen, Synthesis and comparative molecular field analysis (CoMFA) of symmetric and nonsymmetric cyclic sulfamide HIV-1 protease inhibitors, J. Med. Chem. 44 (2001) 155–169.