# The HIV-1 gag and protease coevolution networks

**LI, G.**[1], **VERHEYEN, J.**[2], **RAMON, J.**[3], **EUSÉBIO, M.**[4], **THEYS, K.**[1], **VANDAMME, A.-M.**[1,4]

**1.**Laboratory for Clinical and Epidemiological Virology, Rega Institute for Medical Research, Department of Microbiology and Immunology, KU Leuven, Belgium; **2.**Institute of Virology, University Hospital, University Duisburg-Essen, Essen, Germany **3.**Department of Computer Science, KU Leuven, Leuven, Belgium; **4.**Centro de Malária e Outras Doenças Tropicais, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Portugal

14th EUROPEAN AIDS CONFERENCE
OCTOBER 16–19, 2013
BRUSSELS, BELGIUM

## Objectives

Recent clinical trials reported that some HIV-1 patients who failed the protease inhibitors (PIs) treatment harbor gag mutations. It was hypothesized that gag mutations in the presence or absence of protease mutations may cause PI drug resistance. This study provides further evidence of coevolution between HIV-1 gag and protease under PI drug selective pressure.

## Methodology

We collected 9321 HIV-1 subtype B nucleotide sequence spanning full-length gag and protease from HIV Los Alamos database, together with clinical and experimental drug resistance data published in 60 articles. We created a new assemble-learning coevolution system that integrated 20 known statistical sequence-based methods. HIV-1 gag-protease coevolution networks were trained using 1000 bootstrap estimates and model performance was evaluated based on clinical and experimetnal data and protein contact maps.

## Main Results

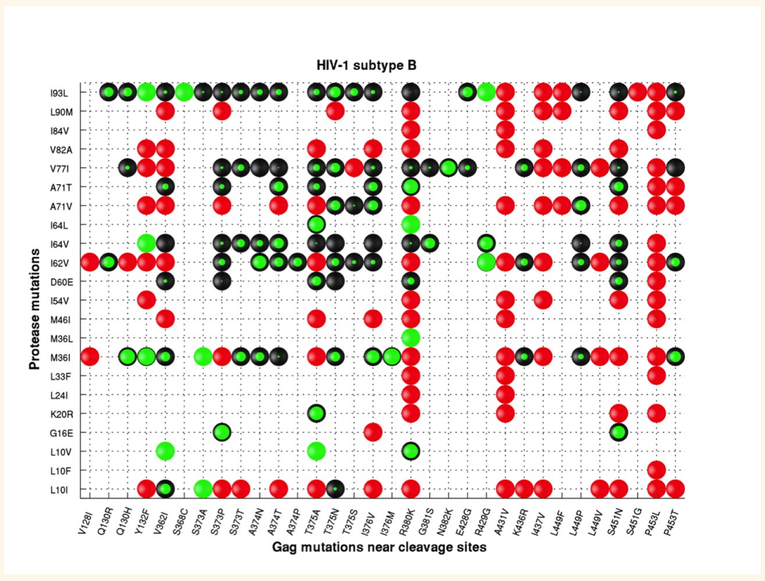


**Figure 1.** Most prevalent CSM-protease mutations in HIV-1 subtype B evaluated by Fisher's exact tests. CSM-protease mutations with significantly different proportions between PI-resistant and PI-resistant+PI-susceptible datasets (p-value $< 0.01$) are indicated with red circles. Other mutations are indicated in black (p-values $> 0.01$), with scaled concentric green circles indicating odds-ratios between 1 and 5; full green circles indicate OR $\geqslant$5.

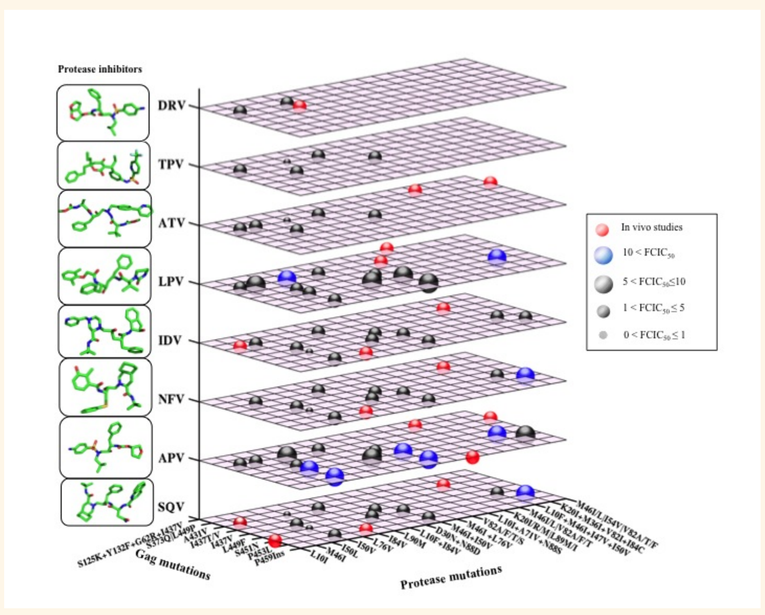


**Figure 2.** Summary of gag-protease mutations observed in in vitro or in vivo experiments. The x-, y- and z-axes represent gag mutations, protease mutations and protease inhibitors, respectively. Crystallized PIs are shown next to the z-axis. Gag-protease mutations were either confirmed in in vitro experiments by log10 fold change (FC) IC50 $> 3$ or in in vivo experiments by association with PI resistance. Degree of drug-specific susceptibility conferred by gag-protease mutations in vitro is represented as: blue (10<FCIC50) spheres and scaled black (large: 5<FCIC50⩽10, medium: 1<FCIC50⩽5, small: 0<FCIC50⩽1). Red spheres show in vivo data.

## Main Results

We discovered that many cleavage sites and positions at the gag C-terminal region between position 370 and 500 were coevolving with protease under PI drug selective pressure. The gag-protease coevolution networks identified 15 gag positions coevolving with 21 positions in protease, of which 10 gag positions were previously associated with PI drug resistance, and all protease positions were included by four HIV-1 genotypic drug resistance interpretation algorithms. All 15 gag positions were located in coiled-coil regions and were outside known HIV-host interaction regions. Gag mutations A431V, I437V, L449F/V, S451G and P453L were strongly associated with protease mutations, while protease mutations L10F, L24I, L33F, M46I, I54V, V82A, I84V and L90M were strongly associated with gag mutations (Fisher's exact tests, FDR-adjusted p-value$<0.01$). A significant difference in the level of co-evolution with protease was observed between the C-terminal region encompassing positions 370 to 500 and the gag region encompassing positions 1 to 369. (p-value$<0.001$).

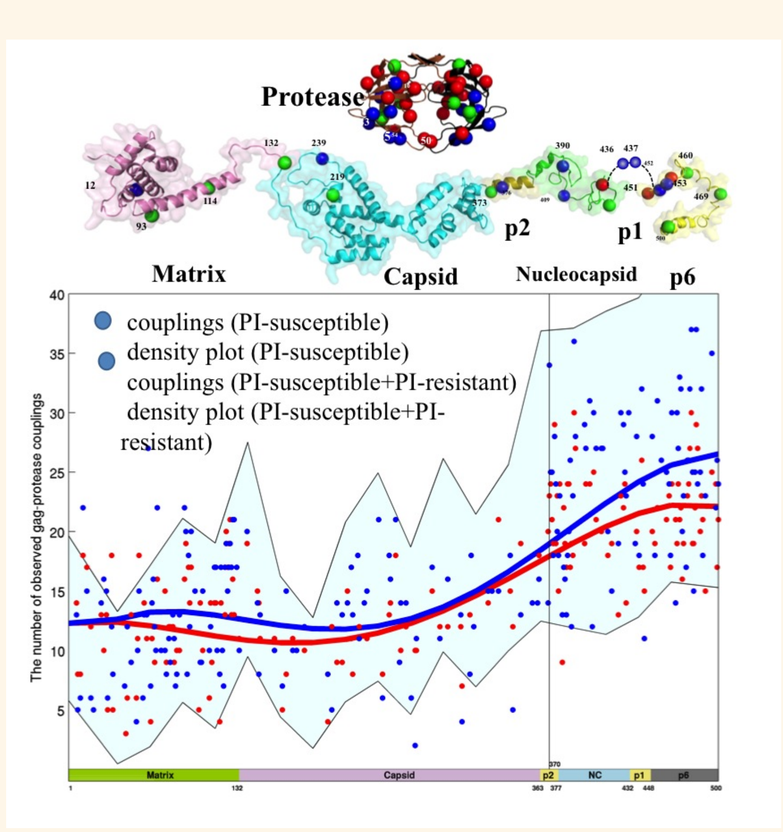


**Figure 3.** (A) Mapping of coevolving positions to HIV-1 protease and gag protein structures (adapted from review Bell, 2013). Spheres are colored red indicating positions identified by both ECS and in vitro or in vivo experiments, green indicating positions identified by ECS but not by in vitro or in vivo experiments, or blue indicating positions identified by in vitro or in vivo experiments but not by ECS. Visualization software: PyMOL v1.5. PDBs: 1HIW, 3H4E, 1TW7, 1U57, 1A1T, 2C55. (B) Distribution of the gag-protease couplings predicted in HIV-1 subtype B full-length gag sequence. X- and y-axes show gag position and number of gag-protease couplings respectively. Dot plot indicates number of gag-protease couplings at each gag position, predicted from PI-susceptible and PI-susceptible+PI-resistant datasets. Polynomial-fitted curves demonstrate the density of gag-protease couplings according to positions.

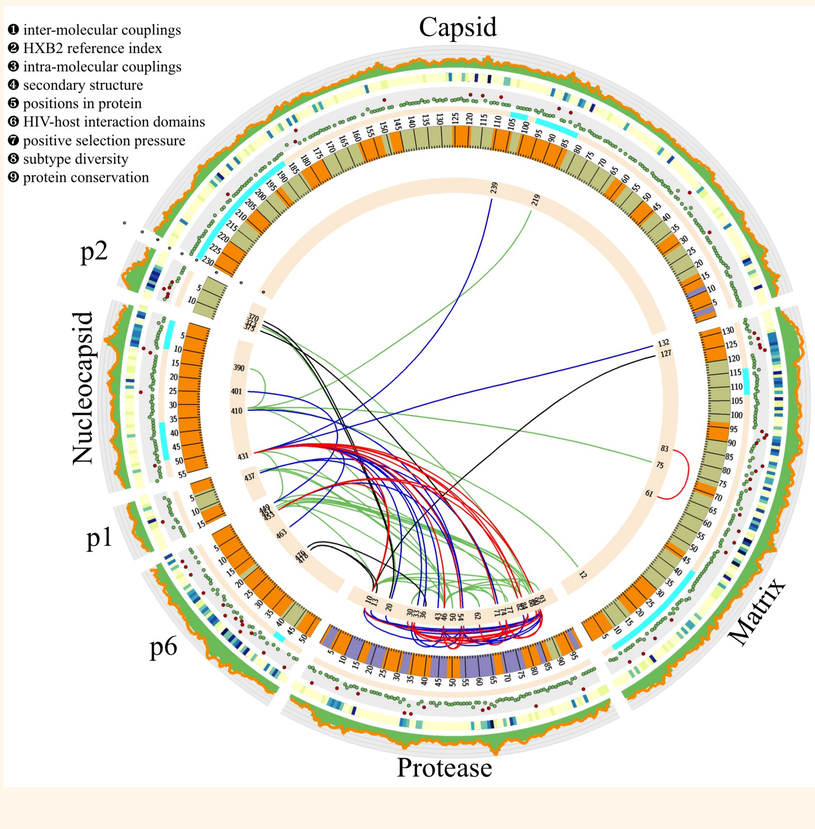


**Figure 4.** The gag-protease coevolution networks. Nine layers are shown: (1) Intermolecular couplings: Red lines indicate ECS-predicted couplings in the set O2-O1 (defined in construction of gag-protease coevolution network in Methods) confirmed by in vitro or in vivo experiments; green lines indicate ECS-predicted couplings in O2-O1 not confirmed by in vitro or in vivo experiments; blue lines indicate couplings confirmed by at least two in vitro or in vivo experiments but not predicted by ECS; black lines indicate the top-ranked couplings in O2 that are not confirmed by in vitro or in vivo experiments. (2) Positions: residue positions indexed according to HXB2 reference. (3) Intramolecular couplings: ECS-predicted intramolecular couplings confirmed by protein contact map (red) and unconfirmed (green). (4) Secondary structures of HIV-1 protease and gag proteins: helix: blue, beta-strand: grey, coiled-coil: orange. (5) Protein positions. (6) HIV-host interaction domains. Interaction domains are colored blue. (7) dN/dS: scatter map of positively selected sites; red dots indicate positions with dN/dS$>1$ and p-values$<0.01$, green indicates others. (8) Inter-subtype diversity: inter-subtype AA diversity illustrated in heatmap (dark blue$>0.1$, light blue:0.001). (9) Protein conservation: conservation scores are plotted in histogram.

## Conclusions

Our data suggest that primarily positions in the gag C-terminal region are coevolving with protease positions associated with drug resistance. We created the first assemble-learning coevolution system in the last two decades, whose superior performance was evaluated by in vivo and in vitro experiments.

## Funding

Collaborative HIV and Anti-HIV Drug Resistance Network (CHAIN, grant Health-F3-2009-223131, European Community's Seventh Framework Programme FP7/2007-2013); BEST HOPE: **B**io-Molecular and **E**pidemiological **S**urveillance of HIV **T**ransmitted Drug Resistance, **H**epatitis Co-Infections and **O**ngoing Transmission **P**atterns in **E**urope (project funded through HIVERA: **H**armonizing **I**ntegrating **V**italizing **E**uropean **R**esearch on HIV/**A**ids, grant 249697); Fonds voor Wetenschappelijk Onderzoek - Flanders (FWO) grant G.0611.09N; FWO K8.012.12N; China Scholarship Council; JSPS funding