Combining molecular dynamic information and an aspherical-atom data bank in the evaluation of the electrostatic interaction energy in multimeric protein-ligand complex: a case study for HIV-1 protease

Prashant Kumar, Paulina M. Dominiak*

Biological and Chemical Research Centre, Department of Chemistry, University of Warsaw,

ul. Żwirki i Wigury 101, Warszawa02-089, Poland, Email: pdomin@chem.uw.edu.pl

Abstract

Computational analysis of protein-ligand interactions is of crucial importance for drug discovery. Assessment of ligand binding energy allows us to have a glimpse on the potential of a small organic molecule to be a ligand to the binding site of a protein target. Available scoring functions such as in docking programs, we could say that they all rely on equations that sum each type of protein-ligand interactions to model the binding affinity. Most of the scoring functions consider electrostatic interactions involving the protein and the ligand. Electrostatic interactions contribute one of the most important part of total interaction energies between macromolecules, unlike dispersion forces they are highly directional and therefore dominate the nature of molecular packing in crystals and in biological complexes and contribute significantly to differences in inhibition strength among related enzyme inhibitors. In this paper, complexes of HIV-1 protease with inhibitor molecules (JE-2147 and Darunavir) have been analysed using charge densities from a transferable aspherical-atom data bank. Moreover, we analyse the electrostatic interaction energy for an ensemble of structures using molecular dynamic simulation to highlight the main features related to the importance of this interaction for binding affinity.

1. Background

The aim of structure based drug design is to predict which potential drug molecules will bind with high affinity and specificity to the target molecule using structural data. Generally, the target molecule is a protein or nucleic acid of known or predicted structure. The target molecule can be a mutated or abnormally produced wild human molecule, an essential component of an infectious agent like a virus, or a toxin. Ideally, the potential drug will be a small, soluble compound. Several new drugs have been developed with designs guided by the crystal structures of proteins, especially for compounds that inhibits the action of enzymes.

HIV protease is a well validated target for AIDS control. Crystal structures of HIV protease with different inhibitors and molecular models were used to guide the design of antiviral drugs for AIDS. However, the present challenge is to overcome the rapid development of drug resistant strains of HIV.[1] Mutations have been observed in more than half of the protease residues on exposure to drugs, and in general, multiple mutations are found with high level resistance.[2] Consequently, the newer drugs must be designed to target different combinations of possible protease mutants. HIV protease is a aspartate protease existing as a homodimer and each monomer contain 99 amino acid residues. The active site lies between the identical subunits and has the characteristic Asp-Thr-Gly (Asp25, Thr26 and Gly27) sequence. The two Asp25 residues (one from each chain) act as the catalytic residues. According to the mechanism for HIV PR protein cleavage, water acts as a nucleophile, which acts in simultaneous conjunction with a well-placed aspartic acid to hydrolyse the scissile peptide bond.[3]

Currently nine HIV-protease inhibitors have been approved by the FDA (indinavir, saquinavir, nelfinavir, ritonavir, amprenavir, lopinavir, atazanavir and tipranavir and darunavir). All these inhibitors can lose their activity when confronted with mutants. Understanding the cause for loss of potency can give new insight to drug designer to develop more improved inhibitors of HIV-protease. There are at least 50 mutation positions identified within the HIV-1 protease.[4–8] From a structural point of view, these mutations have been classified as active site and non-active site depending on whether they are located within or outside the active site cavity.[9] In general, most major mutations occur within the active site and are very conservative, that is, they preserve the charge and polarity and only alter the geometry of the active site. These constraints are dictated by the requirement that the enzyme needs to maintain a sufficient affinity for the substrate and a viable catalytic activity. Because non-active site mutations do not directly affect inhibitor/protease interactions, their effects need to be traced back to a chain of events throughout the protease structure. Originally their role was assumed to be only of a compensatory nature but recent evidence indicates that some of them might play a very important role in lowering the affinity of inhibitors.[10,11]

The estimation of the electrostatic component of the interaction between HIV-1 protease and inhibitors, described in this paper, is based on the transferability of atomic densities, expressed as a superposition of spherical harmonic functions, between atoms in chemically identical environments. It eliminates many of the approximations that are inherent in the commonly used point-charge model in classical force field method. Unfortunately, the point charge model has several limitations. It is unable to take into account such phenomena such as electron polarization, subtle details of electron density anisotropy, and charge penetration. Considerable efforts have been devoted to provide a more realistic description of

intermolecular interactions by including the aforementioned aspects with recent emphasis on penetration energy (Epen).[12–22] The analysis is based on a data bank of transferable atomic densities[23–25], from which the charge density of the protease–inhibitor complexes is reconstructed. Electrostatic interaction energies are then evaluated using an exact algorithm for the short-range interactions and the Buckingham approximation for non-overlapping densities for atoms at large distances[26]. Although electrostatic interaction it is the largest component of the interaction energy, it makes a dominant contribution to the relative orientation of the enzyme and substrate and therefore to molecular recognition and substrate specificity.[27–31]

In this research paper, we report a molecular dynamics (MD) simulation study followed by of a comprehensive electrostatic interactions evaluation for binding of inhibitors to wild type HIV 1 proteases and simultaneously for monomer-monomer interactions in these homodimeric proteins. In addition, we have traced the crucial role of penetration energy in ligand and monomer-monomer binding. We included dynamic component of the complex by computing interaction energies for 100 snapshots from MD and analysing averaged values of energies and their deviations from the average. Aa a result we provide deeper understanding of electrostatic interactions in HIV-1 protein complexes, and propose that introducing the electrostatic component analysis in drug design may allow to construct protocols to find drug molecules effective against the HIV protease resistant variants, too.

2. Theory and computation details

2.1 Molecular Dynamics (MD) Simulation

Molecular dynamics (MD) simulations were performed to simulate the dynamic behaviour of protein in order to reconstruct snapshots of simulated protein electron density and compute pairwise electrostatic energy calculation at atom level which can be further aggregated at residue and chain level. For MD simulation, the crystal structure of the HIV-1 protein (1KZK 1.09 Å and 4DQB 1.5 Å)[32,33] were obtained from the RCSB database. H++ server was used to add hydrogen atoms according to the specified pH of the environment for the protein structure.[34] The HF/6-31G* ab initio level calculations were performed to optimize ligand geometry (JE-2147 and Darunavir respectively) and obtain the electrostatic potential of the ligand using GUASSIAN16 program.[35] To derive the equivalent partial atomic charges for the ligand, RESP[36] fitting was applied on the electrostatic potentials. To assign the atom types, bond, angle, dihedral and van der Waals parameters for the ligand, GAFF force field[37] was used. The Leap module of Amber18 suite[38] was used to prepare the complex for simulation. FF14SB,[39] all atom force fields, was used to parametrize protein atoms. ASP25 was kept protonated.[40–43] Both complexes were neutralized with Cl-/Na+ ion followed by solvating 10 Å buffer of TIP3P[44] water molecules around the system in each direction, forming a truncated octahedral shaped ice cube.

The solvent was first minimized for 1000 steps followed by full system minimization with 4000 steps. Then the system was heated to 298.15 K over 60 ps with a 2 fs step running a further 100 ps MD run for equilibrium. Once an initial equilibrium has been reached, with the temperature- and density-stable, the final stage of the simulation was performed. This consists of running a production simulation at 298.15 K for 2 fs. The system

was then run for 10 ns and stability of trajectory was closely monitored. During simulation, the long-range electrostatic interactions were treated by partial-mesh Ewald method.[45] For short-range non-bonded interactions, 10 Å cut-off was employed. The temperature and pressure were kept constant in the simulation by coupling to the system with Berendsen's thermostat and barostat, respectively.[46] Bond lengths between hydrogens and heavy atoms were constrained using SHAKE.[47] All production-phase simulations were run using GPU accelerated particle-mesh Ewald molecular dynamics (PMEMD) as implemented in Amber18. For the analysis 100 snapshots were extracted from the 10 ns MD trajectory for the system (after stripping all waters and ions).

2.2 Electrostatic interaction energy

To compute the pairwise electrostatic interaction energy (Ees) using UBDB approach, the molecular electron density was represented by Hansen atom model (eq.1) and parameters of the model specific for each atom type were reproduced from the UBDB. The LSDB program was used to transfer the multipole populations from the UBDB to all snapshot structures. This transfer was based on the atomic connectivity and local symmetry recognition. The XDPROP module of the XD2016 package was used to calculate interaction energies from the derived charge density using the Exact Potential Multipole Method (EPMM). The EPMM evaluates the exact Coulomb integral in the inner region (≤4.5 Å) and combines it with a Buckingham-type multipole approximation for long-range interatomic interactions.

$$\rho_k(r) = \underbrace{P_c\rho_c(r_k)}_{\rho_{core}} + \underbrace{P_v\kappa^3\rho_v(\kappa\, r_k)}_{\rho_{valence}} + \underbrace{\sum_{l=0}^{l_{max}} \kappa'^3 R_l(\kappa'r_k) \sum_{m=0}^{l} P_{lm\pm}y_{lm\pm}(\theta, \varphi)}_{\rho_{deformation}} \quad \text{(eq. 1)}$$

3. Results and discussion

### 3.1 HIV-1 Protease interactions with ligand

The trajectory obtained from MD simulation show stable protease complex with 1.815 Å and 1.49 Å average RMS deviations for 1KZK and 4DQB structures, respectively. The low RSMD values suggest good stability of the complexes.

From the MD analysis it was found that two hydrogen bonds stable over the entire simulation were formed between JE-2147 ligand and protein (Fig. 1 top).These are bonds with Asp25A (O21-H13…OD2) and ASP25B (OD2-HD2…O23) residues. There are also hydrogen bonds which alternate with each other. Bonds with Asp30A O, Asp25A OD1 and Asp30A OD1 exists only in the first half of the simulation, and in the remaining half bonds with Ile50A N-H exists in the second half. This is related with alternations of hydrogen bonding between two catalytic residues: Asp25B OD2-HD2 interacts either with OD1 (first half) or OD2 of Asp25A. Hydrogen bonds, and other contacts contributed to total energy of ligand interaction with the protein. The average value of electrostatic interaction energy, computed on the basis of 100 snapshots from MD simulations, amounts to -328(68) kJ/mol for interactions with entire protein, and to -210(42) and -118(46) kJ/mol for interactions with chain A and B, respectively.

Analogously, for darunavir interacting with the protein, 3 hydrogen bonds stable over entire simulation were identified (Fig. 1 bottom). These are bonds with Asp25A OD2, Asp30A O and Asp29B. The remaining identified bonds occur less frequently. There is no alternation in hydrogen bonds present, and only one possible hydrogen bonding between catalytic residues is observed: OD2-HD2 of Asp25B interacting with only OD2 of Asp25A. Entire darunavir molecule interacts with entire protein with electrostatic energy of -283(52) kJ/mol, and with

particular monomers with Ees of -184(44) and -99(35) kJ/mol for monomer A and B, respectively.
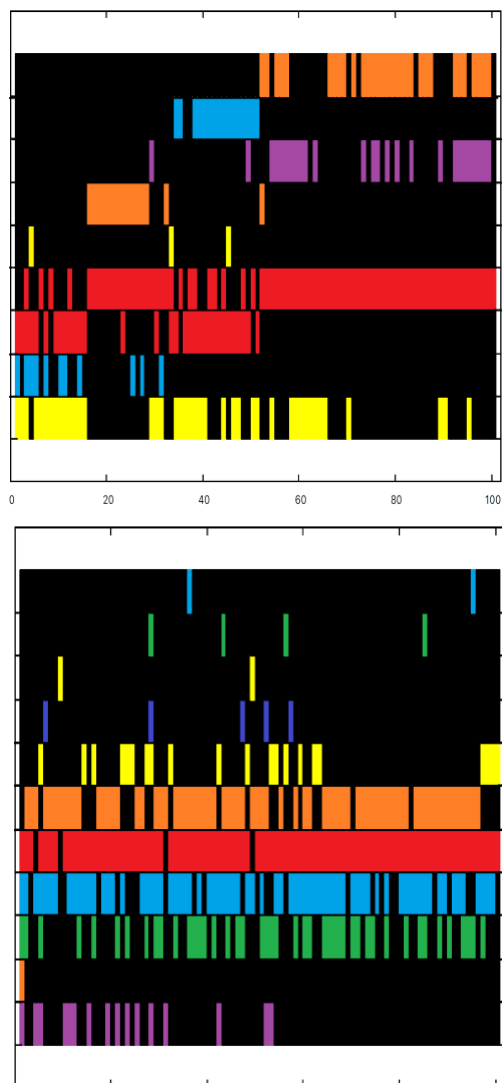
Donor		Acceptor		D…A	D-H…A	Freq.
ASP25B	OD2-HD2	ASP25A	OD2			
JE-2147	O2-H6	ASP30A	OD1	2.70	161.0	0.16
ILE50A	N-H	JE-2147	O10	2.85	160.1	0.26
ASP25B	OD2-HD2	ASP25A	OD1			
ASP25B	OD2-HD2	JE-2147	O21	2.78	145.0	0.03
JE-2147	O21-H13	ASP25A	OD2	2.66	161.1	0.79
JE-2147	O21-H13	ASP25A	OD1	2.68	155.4	0.32
JE-2147	O2-H6	ASP30A	O	2.81	144.6	0.11
ASP25B	OD2-HD2	JE-2147	O23	2.71	163.8	0.42

Donor		Acceptor		D…A	D-H…A	Freq.
ASP30A	N-H	DRV	N1	2.96	157.6	0.02
ASP29B	N-H	DRV	O26	2.91	142.2	0.04
DRV	O18-H18	ASP25B	OD1	2.68	156.8	0.02
ASP30B	N-H	DRV	O26	2.94	162.5	0.05
ASP25B	OD2-HD2	DRV	O18	2.73	167.8	0.21
ASP25B	OD2-HD2	ASP25A	OD2			
DRV	O18-H18	ASP25A	OD2	2.66	166.4	0.96
DRV	N1-H	ASP30A	O	2.85	159.3	0.74
ASP29B	N-H	DRV	O28	2.87	161.9	0.45
ASP25B	OD2-HD2	ASP25A	OD1			
ILE50	N-H	DRV	O10	2.91	155.9	0.16

Figure 1. Hydrogen bonds identified in 100 snapshots of MD simulations for 1KZK (top) and 1DQB (bottom) structures including average values of donor…acceptor distances (A) and hydrogen bond angles (º), along with the graphical representation of their occurrence (coloured bars) in particular snapshot.

3.2 Monomer-monomer interactions in HIV-1 Protease

The strength of ligand protein interactions can be further contrasted with the strength of interactions between monomer A and monomer B of protein. Again the analysis is based on averaged values of energies computed for 100 snapshots from MD simulations. It appears that the strongest electrostatic attractive or repulsive interactions between a residue from one monomer and all residues of the other monomer are of similar magnitude as interactions of a ligand with a monomer, Fig. 2. Moreover, it is remarkable how interaction profiles are similar for 1KZK and 1DQB structures, despite the facts that the MD simulations started from different crystallographic structures, structures contains different ligands, and sequences of proteins are not exactly the same. The proteins come from different subtypes of HIV-1 virus (group M), from subtype B or D for 1KZK and 1DQB structures, respectively, which differ as following: Leu33→Ile, Pro63→Ile, Val64→Ile ,Cys67→Ala, Cys95→Ala.

The strongest attractive interactions with entire second monomer are observed for residue Asp25 from monomer A, and for Asp29, Asp30 and Phe99 residues from both monomers. These are residues which contribute to the constitution of the active site of the protease (Asp25, Asp29, Asp30), including one of the catalytic residues (Asp25 A), or belongs to the most interpenetrating fragment of a dimer interface, where two β-strands from one monomer and two β-strands from the other one form antiparallel terminal β-sheet.

Closer look into electrostatic interactions between particular residues (not a residue and entire monomer) reveals that Asp29 residues forms the strongest residue-residue interactions, Fig 3a. Asp29 from monomer B interacts with Arg8 from monomer A with the most negative Ees, -186(30) kJ/mol for 1KZK and -220(23) kJ/mol for 1DQB, and the pair is

9

followed by its symmetrical counterpart, Asp29A…Arg8B, Ees = -160(70) and -202(34) kJ/mol. Residues Arg8 (B and A) are also the ones which exhibit the largest repulsive interactions with a residue from the other monomer, they interact with Arg87 (A and B) with Ees equal to 64(1) and 55(4) kJ/mol, and to 65(3) and 64(5) kJ/mol for 1KZK and 1DQB, respectively. Interestingly, Asp25 from monomer A (deprotonated) is not interacting particularly strongly with one individual residue, but rather its interactions with many residues are somewhat similar and sum-up to large total negative value of Ees. Ees energies for Asp25 from monomer B (protonated), on the other hand, is not summing up to such large value, most probably because it is not charged anymore, and its long range electrostatic interactions are weaker. Like Asp25 A, Asp30 residues interact similarly strongly with many residues from the second monomer. Phe99 residues, on the other hand, interact particularly strongly with Pro1 residues, with Ees in the range of –116 to -170 kJ/mol. The last strongly interacting pairs of residues are Asp98 interacting with Thr96, with Ees in the range of -64 to -100 kJ/mol.
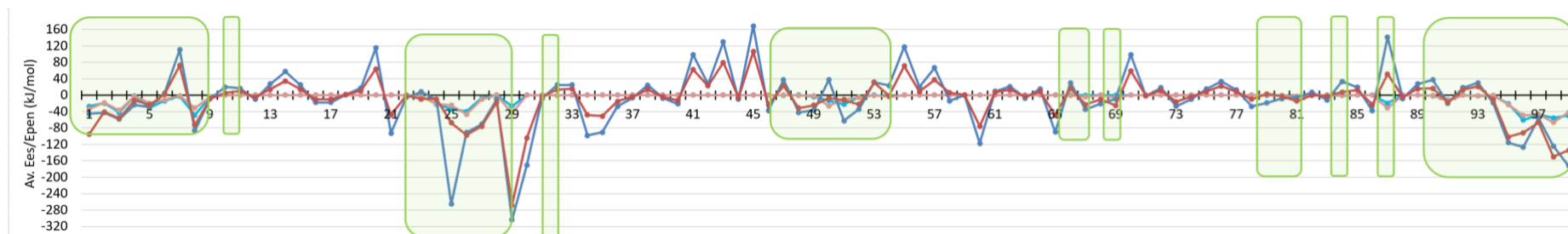
(a)
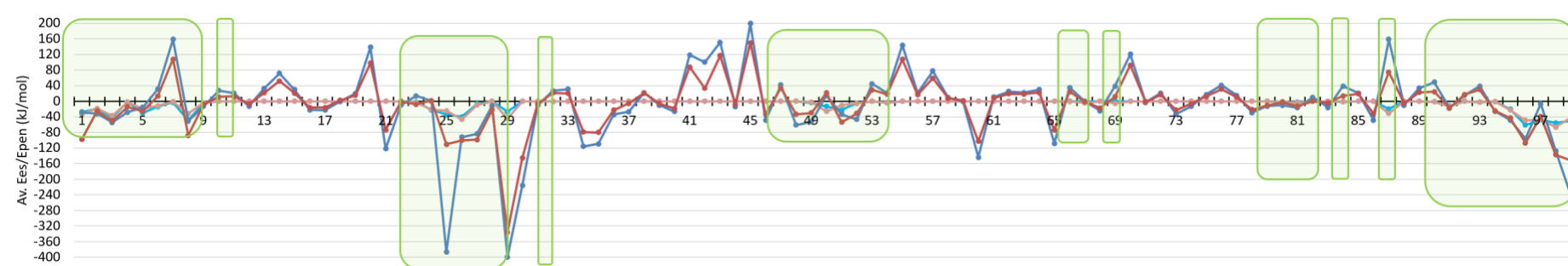
(b)

Figure 2: Averaged electrostatic interaction energies (Ees) and penetration contributions to them (Epen) for interactions of a residue from one monomer with all residues of the other monomer. Ees: dark blue – monomer A, dark red – monomer B; Epen: light blue – monomer A, light red – monomer B. In green regions of Epen > 0.0 kJ/mol are highlighted.

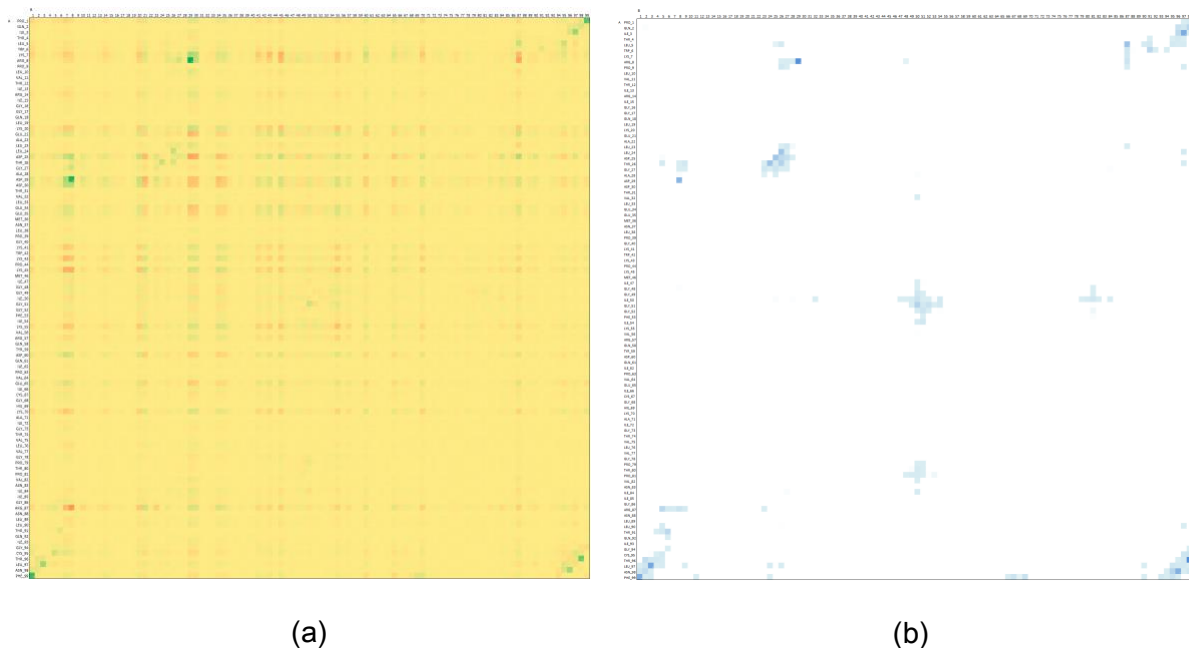(a)   (b)

Figure 3: Heat plot for averaged electrostatic interaction energies (a) and penetration contributions to them (b) for interactions of a residue from monomer A (rows) with a residue of the monomer B (columns) for 1KZK. Scale for (a): from green to red, dark green – the most negative values, dark red – the most positive values, yellow – close to zero values. Scale for (b): from blue to white, dark blue – the most positive values, white – zero values.
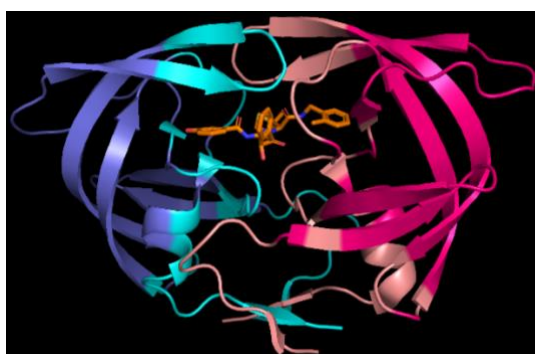

Figure 4: 1KZK structure with residues exhibiting non-zero penetration energies showed in light blue (monomer A) and light red (monomer B).

All the above mentioned residues, but Asp30, besides showing large negative or positive values of Ees are characterized by significant values of penetration contributions, Epen, to electrostatics interactions. It means their interactions should be considered as short-range, because at least one atom from one residue interpenetrate electron density of the other atom(s) from the second residue. Epen is particularly helpful in identifying pairs of interacting molecular fragments which are interacting with each other directly (Fig. 3b, Fig. 4). It is much better than any other simplified method of distance characterisation like the one showed on Fig. S1(c). Non-zero Epen, in addition, indicate that other contributions to total interaction energies like exchange-repulsion, dispersion and induction, will be most probably not negligible for that interaction.

Here Epen allows easily to identify not only regions of active site (residues 23-29) built by mutual interactions of both monomers and regions of N- and C-termini interacting to lock the dimer (residues 1-9 and 90-99), as mentioned above, but also other well-known region of HIV-1 protease, namely two flaps covering the active site. Residues from 47 to 54 of each monomer directly interact with each other, but their electrostatic interactions are not as strong as in the other mentioned regions. It correspond very well with the loops ability to open spontaneously, what plays a crucial role in the mechanism of substrate binding. Also Epen shows that there are direct interactions between the active site and terminal β-sheet fragments of dimer interface through residues Thr26, Gly27 and Asp29 from one monomer interacting with residues Thr4, Lys7 and Pro9 from the other monomer.

The remaining interacting residues exhibiting larger than zero Epen are the following: Val11, Val32, Ile66, Cys67, His69, Pro79-Thr80-Pro81-Val82, Ile84, Arg87. Residues Arg87

was already mentioned in the contexts of strong repulsive interactions with Arg8 from the terminal β-sheet dimer interface. The rest do not interact strongly electrostically, but nevertheless they complete constitution of dimer interface.

Interestingly, there are many residues with zero charge penetrations, meaning interacting true long-range electrostatic interactions only, showing quite large values of positive or negative energy, ca. +/- 80 kJ/mol. Residues acting towards monomer-monomer repulsion are all positively charged: Arg14, Lys20, Lys41, Lys43, Lys45, Lys55, Arg57, Lys70. Residues strengthening monomer-monomer interactions through long-range electrostatic attractions are those negatively charged: Glu21. Glu34, Glu35, Asp60, Glu65.

All the above analyses were done on the basis of energies averaged for 100 snapshots from MD. We believe they give more accurate view on interaction energies, as they are not bias by single realizations of one state from many equally possible, if existing. Such analysis also shows that that there are residues, which exhibit large fluctuations in Ees, but small in overall atom-atom distances: Asp25, Asp29, Ile50 and Arg87, and slightly less visible Thr26 and Asp30. Here rotation of single functional group must influence a lot electrostatic interactions, like –COO- from Asp29 interacting with guanidine group from Arg8, for which co-planar orientation is the most optimal, or –COOH group from Asp25B forming O-H..O hydrogen bond with Asp25A instead with the ligand.

3.3 Mutants disrupting monomer-monomer interactions in HIV-1 Protease

Among many mutants observed in HIV-1 protease, there three often mentioned in the context of weakening of monomer-monomer interactions: Thr26→Ala, Asp29→Asn and Arg87→Lys. Interestingly, all these mutants are among the residues which exhibit large Ees fluctuations but relatively small movements of most of the atoms. The mutation of the remaining Asp30 residue is also observed but mention (Asp30→Asn) in the context of ligand binding. Asp29 and Arg87 are also among the ones most strongly interacting electrostatically.

Electrostatic intereaction between protein-protein and protein-ligand using apshreical atom databank (UBDB) over snapshots from trajectory obtained from molecular dynamic simulation. The electrostatic interaction is a non-negligible component of the interaction energy driving the strength of HIV-1 protease inhibitors and must be embraced when designing new inhibitors. The model applied in our work takes into account the asphericity of the atoms as atom types in the UBDB. Unlike point-charge models, it takes into account for the directionality of the atom–atom interactions. The model is applicable to microscopic analysis of structure–function activity in biological molecules. Moreover, the electrostatic interaction is much stronger than vdW interaction, it can provide sufficient binding strength without the necessity of increasing the bulkiness of the designed molecules and itself can be used a sole scoring function. Therefore the new designed molecules will be not only effective against wild type protease but also against its drug resistant mutant variants. In the next step of the research work, we would like to give more emphasis on role of electrostatic interaction in various kind of protease mutants in order to understand the drug specificity against mutants.

4. Conclusion and outlook

Electrostatic interactions between protein monomer-monomer and protein-ligand were analysed using aspherical atom databank (UBDB) over snapshots from trajectory obtained from molecular dynamic simulation. The electrostatic interaction is a non-negligible component of the interaction energy driving the strength of HIV-1 protease inhibition and must be embraced when designing new inhibitors. The model applied in our work takes into account the asphericity of atoms by using aspherical electron density fragments specific for particular atom types stored in the UBDB. Unlike point-charge models, the UBDB approach takes into account the directionality of the atom–atom interactions and charge penetration. The model is applicable to microscopic analysis of structure–function activity in biological molecules. Moreover, the electrostatic interaction is much stronger than vdW interaction, it can provide sufficient binding strength without the necessity of increasing the bulkiness of the designed molecules and itself can be used a sole scoring function. Expansion of protein-ligand interaction analyses by analysis of monomer-monomer interactions in multimeric proteins gives wider contexts for protein ligand interactions. And together with inclusion of dynamic aspects of protein-ligand complex helps to understand where are the hot spots important for dimerization of possible mutations. Therefore the new designed molecules will be not only effective against wild type protease but also against its drug resistant mutant variants. In the next step of the research work, we would like to give more emphasis on role of electrostatic interaction in various kind of protease mutants in order to understand the drug specificity against mutants.

5. Funding information

References

1. Appadurai, R. & Senapati, S. Dynamical Network of HIV - 1 Protease Mutants Reveals the Mechanism of Drug Resistance and Unhindered Activity. (2016). doi:10.1021/acs.biochem.5b00946

2. Shafer, R. W. Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clin. Microbiol. Rev.* **15**, 247–277 (2002).

3. Jaskólski, M. *et al.* Structure at 2.5-A resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry* **30**, 1600–1609 (1991).

4. Gulnik, S., Erickson, J. W. & Xie, D. HIV protease: enzyme function and drug resistance. *Vitam. Horm.* **58**, 213–256 (2000).

5. Kozal, M. J. *et al.* Extensive polymorphisms observed in HIV–1 clade B protease gene using high–density oligonucleotide arrays. *Nat. Med.* **2**, 753–759 (1996).

6. Shafer, R. W., Chuang, T. K., Hsu, P., White, C. B. & Katzenstein, D. A. Sequence and drug susceptibility of subtype C protease from human immunodeficiency virus type 1 seroconverters in Zimbabwe. *AIDS Res. Hum. Retroviruses* **15**, 65–69 (1999).

7. Boden, D. & Markowitz, M. Resistance to human immunodeficiency virus type 1 protease inhibitors. *Antimicrob. Agents Chemother.* **42**, 2775–2783 (1998).

8. Hertogs, K. *et al.* Phenotypic and genotypic analysis of clinical HIV-1 isolates reveals extensive protease inhibitor cross-resistance: a survey of over 6000 samples. *AIDS* **14**, 1203–1210 (2000).

9. Erickson, J. W. & Burt, S. K. Structural mechanisms of HIV drug resistance. *Annu. Rev. Pharmacol. Toxicol.* **36**, 545–571 (1996).

10. Muzammil, S., Ross, P. & Freire, E. A major role for a set of non-active site mutations in the development of HIV-1 protease drug resistance. *Biochemistry* **42**, 631–638 (2003).

11. Olsen, D. B. *et al.* Non-active site changes elicit broad-based cross-resistance of the HIV-1 protease to inhibitors. *J. Biol. Chem.* **274**, 23699–23701 (1999).

12. Bojarowski, S. A., Kumar, P. & Dominiak, P. M. A Universal and Straightforward Approach to Include Penetration Effects in Electrostatic Interaction Energy Estimation. *ChemPhysChem* **17**, 2455–2460 (2016).

13. Kumar, P. *et al.* A comparative study of transferable aspherical pseudoatom databank and classical force fields for predicting electrostatic interactions in molecular dimers. *J. Chem. Theory Comput.* **10**, 1652–1664 (2014).

14. Bojarowski, S. A., Kumar, P., Wandtke, C. M., Dittrich, B. & Dominiak, P. M. Universal Method for Electrostatic Interaction Energies Estimation with Charge Penetration and Easily Attainable Point Charges. *J. Chem. Theory Comput.* acs.jctc.8b00781 (2018). doi:10.1021/acs.jctc.8b00781

15. Kramer, C., Spinn, A. & Liedl, K. R. Charge anisotropy: Where atomic multipoles matter most. *J. Chem. Theory Comput.* **10**, 4488–4496 (2014).

16. Shi, Y. *et al.* Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory Comput.* **9**, 4046–4063 (2013).

17. Söderhjelm, P. & Ryde, U. How accurate can a force field become? A polarizable multipole model combined with fragment-wise quantum-mechanical calculations. *J. Phys. Chem. A* **113**, 617–627 (2009).

18. Stone, A. J. Electrostatic damping functions and the penetration energy. *J. Phys. Chem. A* **115**, 7017–7027 (2011).

19. Wang, B. & Truhlar, D. G. Screened Electrostatic Interactions in Molecular Mechanics. *J. Chem. Theory Comput.* (2014). doi:10.1021/ct5005142

20. Turner, M. J., Grabowsky, S., Jayatilaka, D. & Spackman, M. A. Accurate and Efficient Model Energies for Exploring Intermolecular Interactions in Molecular Crystals. *J. Phys. Chem. Lett.* (2014).

21. Bojarowski, S. A., Kumar, P. & Dominiak, P. M. Interplay of point multipole moments and charge penetration for intermolecular electrostatic interaction energies from the University at Buffalo pseudoatom databank model of electron density. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **73**, 598–609 (2017).

22. Spackman, M. a. The use of the promolecular charge density to approximate the penetration contribution to intermolecular electrostatic energies. *Chem. Phys. Lett.* **418**, 158–162 (2006).

23. Dominiak, P. M., Volkov, A., Messerschmidt, M., Li, X. & Coppens, P. A Theoretical Databank of Transferable Aspherical Atoms and Its Application to Electrostatic Interaction Energy Calculations of Macromolecules. *J. Chem. Theory Comput.* **3**, 232-

247. (2007).

24. Jarzembska, K. N. & Dominiak, P. M. New version of the theoretical databank of transferable aspherical pseudoatoms, UBDB2011 - Towards nucleic acid modelling. *Acta Crystallogr. Sect. A Found. Crystallogr.* **68**, 139–147 (2012).

25. Kumar, P., Gruza, B., Bojarowski, S. A. & Dominiak, P. M. Extension of the transferable aspherical pseudoatom data bank for the comparison of molecular electrostatic potentials in structure{--}activity studies. *Acta Crystallogr. Sect. A* **75**, 398–408 (2019).

26. Volkov, A., Li, X., Koritsanszky, T. & Coppens, P. Ab Initio Quality Electrostatic Atomic and Molecular Properties Including Intermolecular Energies from a Transferable Theoretical Pseudoatom Databank. *J. Phys. Chem. A* **108**, 167–180 (2004).

27. Boggavarapu, R., Jeckelmann, J.-M., Harder, D., Ucurum, Z. & Fotiadis, D. Role of electrostatic interactions for ligand recognition and specificity of peptide transporters. *BMC Biol.* **13**, 58 (2015).

28. Sartori, G. R. & Nascimento, A. S. Comparative Analysis of Electrostatic Models for Ligand Docking. *Front. Mol. Biosci.* **6**, 52 (2019).

29. Bitencourt-Ferreira, G., Veit-Acosta, M. & de Azevedo, W. F. J. Electrostatic Energy in Protein-Ligand Complexes. *Methods Mol. Biol.* **2053**, 67–77 (2019).

30. Masukawa, K. M., Kollman, P. A. & Kuntz, I. D. Investigation of Neuraminidase-Substrate Recognition Using Molecular Dynamics and Free Energy Calculations. *J. Med. Chem.* **46**, 5628–5637 (2003).

31. Ahsan, M., Pindi, C. & Senapati, S. Electrostatics Plays a Crucial Role in HIV-1 Protease Substrate Binding, Drugs Fail to Take Advantage. *Biochemistry* **59**, 3316–3331 (2020).

32. Reiling, K. K., Endres, N. F., Dauber, D. S., Craik, C. S. & Stroud, R. M. Anisotropic Dynamics of the JE-2147−HIV Protease Complex: Drug Resistance and Thermodynamic Binding Mode Examined in a 1.09 Å Structure,. *Biochemistry* **41**, 4582–4594 (2002).

33. Mittal, S., Cai, Y., Nalam, M. N. L., Bolon, D. N. A. & Schiffer, C. A. Hydrophobic Core Flexibility Modulates Enzyme Activity in HIV-1 Protease. *J. Am. Chem. Soc.* **134**, 4163–4168 (2012).

34. Gordon, J. C. *et al.* H++: a server for estimating p Ka s and adding missing hydrogens to macromolecules . *Nucleic Acids Res.* **33**, W368–W371 (2005).

35. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. *et al.* Gaussian 09, Revision D.01. (2009).

36. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A Well-Behaved Electrostatic

Potential Based Method Using Charge Restraints for Deriving. *J. Phys. Chem.* **97**, 10269–10280 (1993).

37. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).

38. D.A. Case et al, AMBER 2018, University of California, S. F. Amber. (2018).

39. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* **78**, 1950–1958 (2010).

40. Smith, R., Brereton, I. M., Chai, R. Y. & Kent, S. B. Ionization states of the catalytic residues in HIV-1 protease. *Nat. Struct. Biol.* **3**, 946–950 (1996).

41. Torbeev, V. Y. & Kent, S. B. H. Ionization state of the catalytic dyad Asp25/25' in the HIV-1 protease: NMR studies of site-specifically 13C labelled HIV-1 protease prepared by total chemical synthesis. *Org. Biomol. Chem.* **10**, 5887–5891 (2012).

42. Adachi, M. *et al.* Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. *Proc. Natl. Acad. Sci.* **106**, 4641–4646 (2009).

43. Wang, Y.-X. *et al.* Solution NMR Evidence That the HIV-1 Protease Catalytic Aspartyl Groups Have Different Ionization States in the Complex Formed with the Asymmetric Drug KNI-272. *Biochemistry* **35**, 9945–9950 (1996).

44. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

45. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

46. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

47. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).

Supporting Information
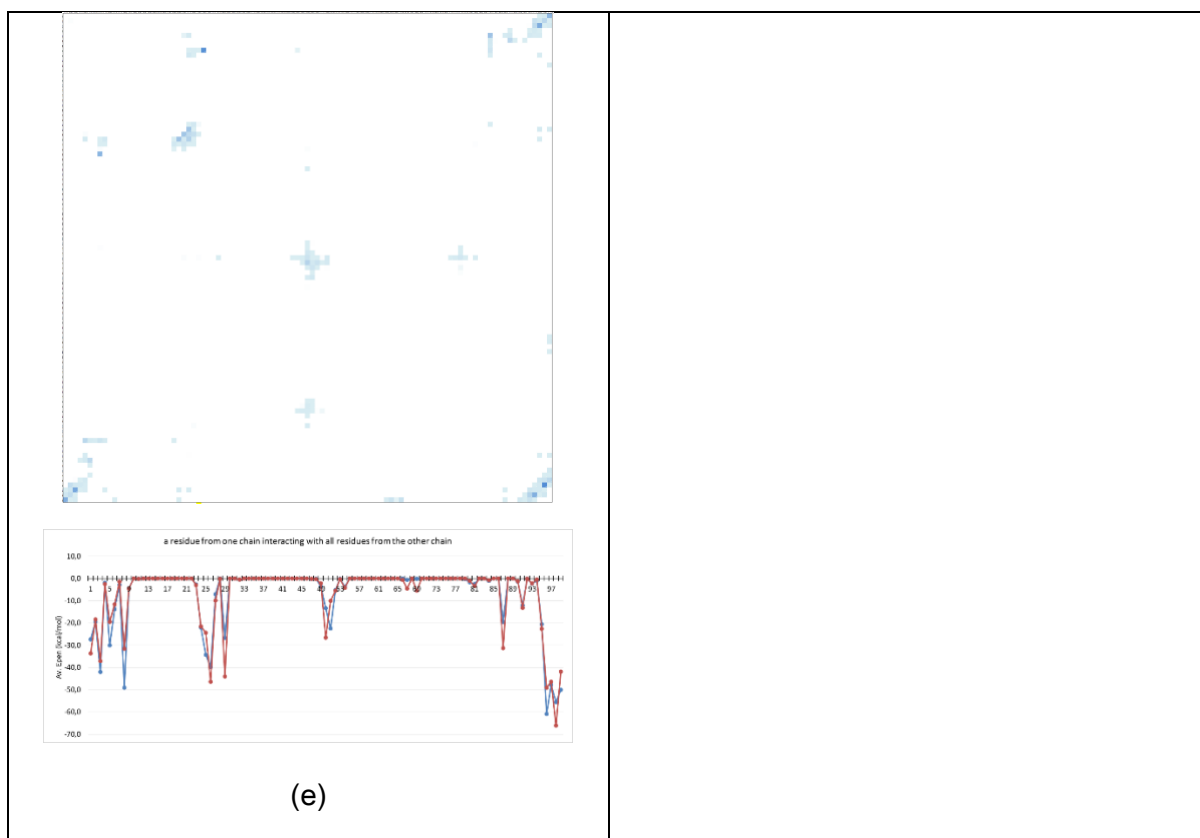

(a)

(b)

(c)

(b)

(e)
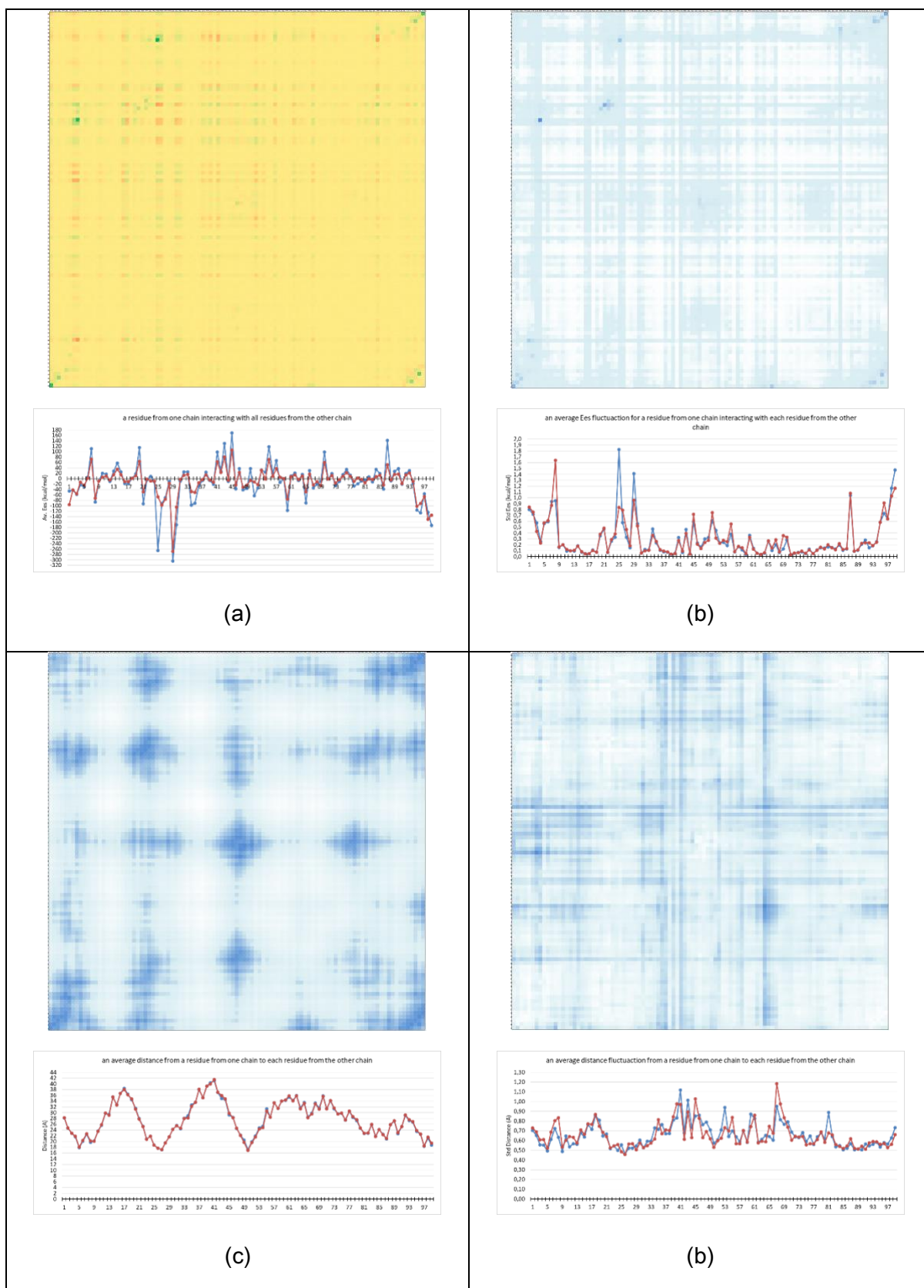
Figure S1: Heat plots and 1D profiles for averaged electrostatic interaction energies (a), standard deviations of electrostatic interaction energies (b), averaged atom-atom distances summed per residue pair (c); standard deviations of atom-atom distances summed per residue pair (d), and penetration contributions to electrostatic energies (e). For detailed descriptions see Figures 2 and 3.