Supporting Information

Allosteric communication networks in proteins revealed through pocket crosstalk analysis

Giuseppina La Sala^{1,§}, Sergio Decherchi^{2,3,§}, Marco De Vivo^{1,4,*} and Walter Rocchia^{2,*}

1. Laboratory of Molecular Modeling and Drug Discovery, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy

2. CONCEPT Lab, Istituto Italiano di Tecnologia,Via Morego 30, 16163 Genova, Italy

3. BiKi Technologies s.r.l., via XX Settembre 33, 16161 Genova, Italy

4. IAS-S/INM-9 Computational Biomedicine Forschungszentrum Jülich Wilhelm-Johnen-Staße 52428 Jülich, Germany

§ These authors contributed equally

*Corresponding authors:

Dr. Marco De Vivo

Phone: +39 01071781577

Marco.devivo@iit.it

Dr. Walter Rocchia

Phone: +39 01071781552

Walter.rocchia@iit.it

1. Systems preparation

All the systems simulated in the present work were prepared by means of AmberTools, version $10¹$ and Gaussian² for the ligand partial charge computation.

PNP. The PNP system consists of PNP homotrimeric unit with 9 DADME-ImmH ligand and phosphate ions. The PNP trimer was modeled starting from the PDB code 3K8O using the chains E, Q and Y. The missing first residue (glutammate) in chains E and Q was added manually after superimposing the two chains with chain Y. The selected tautomerization of H257 was epsilon according to Hirschi et al.³ DADME-ImmH ligand was retrieved from PDB code 1RSZ and placed around the PNP trimer, while phosphate ions were placed in the same positions found in PDB code 1RR6 X-ray structure.⁴ Partial charges of DADME-ImmH and phosphate ions were computed at the HF/6-31G* theory level, imposing a net charge of +1 for DADME-ImmH. General Amber Force Field $(GAFF)^5$ was employed to parameterize the ligands and the ions, while partial charges were fitted using the RESP procedure (via Antechamber). $6\degree$ O2-P-OH and HO-OH-P angles were modified in order to avoid failures of the SHAKE algorithm.⁷ Finally, protein was parameterized with AmberFF99SB-ILDN force field⁸ and immersed in TIP3P water⁹ box of ~11,000,000 \AA^3 , containing nearly 100,000 atoms in total.

Adenosinic Receptor (A2A). The A2A system was built using the PDB code 3UZC X-ray structure.¹⁰ The missing loop (residues 150-157) was modeled using PDB code $4E[Y]$ ¹¹ as template. Alanine was mutated to serine to reproduce the wild type system. The protein was embedded in a POPC membrane bilayer of 75 x 75 \AA ² size with a water layer⁹ of 30 \AA for each side. The membrane was assembled with CharmmGUI server¹² while the posing and the physical plugging of the protein within the bilayer was done with Membrane Tool present in BiKi Life Sciences Software Suite.

Abl kinase. For Abl kinase we defined 5 model systems, namely a) KD_{out} , b) KD_{in} , c) T315I-KD_{in}, d) Myr/KD_{in} and e) Myr/T315I-KD_{in}. System *a* was built from 1OPL chain B X-ray structure,¹³ after the removal of SH2 domain (see ref. 14 for a comprehensive description of the modeling procedures). Conversely, systems *b-e* were modeled starting from the 2F4J X-ray structure.¹⁵ Both

T315I-KD_{in} and Myr/T315I-KD_{in} were build starting from KD_{in} and replacing the gatekeeper T315 with isoleucine. In addition, in both $Myr/T315I-KD_{in}$ and Myr/KD_{in} the myristate molecule was manually added using as reference structure reference the PDB code 1OPL chain A. Point charges of myristate were computed at the HF/6-31G* level theory, while the General Amber Force Field $(GAFF)^5$ was used to parameterize the myristate. All the systems were parameterized with Amber99SB force field¹⁶ and immersed in TIP3P water⁹ box.

2. Molecular dynamics (MD) simulations

Pnp. The system was first equilibrated for 350 ps in NVT ensemble in 5 steps: 150 ps constraining all protein heavy atoms using a harmonic constant 40 kcal mol⁻¹ \AA^{-2} ; 50 ps constraining the protein backbone using a harmonic constant 20 kcal mol⁻¹ \AA ⁻²; 50 ps where the protein backbone is harmonically constrained with a constant of 10 kcal mol⁻¹ \AA ⁻² and 50 ps with a constant of 1 kcal mol⁻¹ A^{-2} for the protein backbone. Then the system underwent to other 350 ps of MD simulations in NPT ensemble at 1 bar. For the production phase we run about 1000 ns of simulations in NVT ensemble using the ACEMD engine.¹⁷ The equation of motion was integrated every 2 fs. Bond involving hydrogen atoms were restrained to their equilibrium length with the SHAKE algorithm.¹⁸ A short-range nonbonded cutoff of 9 Å was applied, whereas long-range electrostatics was treated with the Particle Mesh Ewald method.¹⁹ Temperature was maintained at 300 K using the Langevin thermostat²⁰ and damping of 0.1 ps.

Adenosinic Receptor (A2A). The system was equilibrated for 300 ps in the NVT ensemble employing 3 steps of 100 ps each at 100, 200 and 300 K. In the first two steps a harmonic constraint of 1000 kJ/mol \AA^2 was applied to the protein backbone. Then, 1 ns in the NPT ensemble was performed to reach pressure equilibrium. The production was run for 100 ns in the NPT ensemble. A short-range nonbonded cutoff of 11 Å was applied, whereas long-range electrostatics was treated with the PME method.¹⁹ The Bussi-Parrinello thermostat²¹ was employed. Pressure was kept anisotropic (X,Y plane). Analogous settings were used for the adiabatic bias simulations; here the first frame of plain production MD was used. MD simulations were performed with the Gromacs 4.6.1 engine²² and Plumed2²³ for the enhanced sampling.

Abl kinase. First the systems underwent to a heating phase, in which the systems reached 300 K in 100 ps in the NVT ensemble, keeping the Cα atoms fixed in their original positions, using an harmonic force constant of 1000 kJ/mol \AA^2 . Then the pressure was equilibrated running 5 ns of NPT simulation reaching 1 bar and maintaining the Ca restrained. Production phase was performed in the NPT ensemble. Bond involving hydrogen atoms were restrained with the LINCS algorithm.²⁴ A short-range nonbonded cutoff of 9 Å was applied, whereas long-range electrostatics was treated with the PME method.¹⁹ The equations of motion were integrated every 2 fs. The Parrinello-Rahaman barostat²⁵ and the velocity-rescaling thermostat²¹ were employed with a relaxation time τ of 2 ps and 0.1 ps, respectively. Coordinates of the systems were collected every 50 ps for each run. All the simulations were performed with the Gromacs 4.6 engine.²² Overall, we collected \sim 1.8 µs for SH2-KD_{wt}, ~1.5 µs for SH2-KD_{T231R}, ~1.3 µs for SH2-KD_{I164E}, ~1.3 µs for KD_{in}, ~1.3 µs for KD_{out}, ~1.0 µs for T315I-KD_{in}, ~0.8 µs for Myr/T315I-KD_{in} and ~1.0 µs for Myr/KD_{in}.

3. Dasatinib residence time estimation

Here, we employed scaled MD simulations to compare the residence time of dasatinib in two molecular systems: i) the T315I form of Abl KD (Das/T315I-KDin); and ii) the T315I form of Abl KD in complex with the myristate molecule (Das-Myr/T315I-KD_{in}). Both complexes were prepared starting from T315I-KD_{in} and Myr/T315I-KD_{in} molecular systems, previously described. The dasatinib was manually placed within the ATP binding site after superposition with the 2GQG Xray-structure of wild-type Abl. As reported before for the myristate, dasatinib was parameterized using the General Amber Force Field $(GAFF)^5$, while point charges were computed at HF/6-31G* level theory. For both Das/T315I-KD_{in} and Das-Myr/T315I-KD_{in} systems, a plain MD simulation was run to remove the steric clashes between the dasatinib and the ATP pocket and relax the starting complexes. In detail, after a minimization stage, the systems were heated at 300 K using the velocity-rescaling thermostat²¹ and brought at 1 bar using the Parrinello-Rahman barostat²⁵. Subsequently, a 100 ns-long production stage was conducted in the NVT ensemble. At this point we performed on both trajectories a cluster analysis, using the BiKi LifeSciences clustering algorithm.⁷ For each system, we retrieved the medoid structure of the most populated cluster, which can be reasonably considered the most representative conformation of the MD simulations. The resulting structures were subsequently employed as starting points for a series of scaled-MD simulations.^{26, 27} For both Das/T315I-KD_{in} and Das-Myr/T315I-KD_{in} systems, 20 scaled-MD simulations were performed, using a smoothing coefficient $\lambda = 0.4$. To prevent protein unfolding, we applied a weak restraint (50 kJ mol⁻¹ nm⁻¹) to the whole backbone with exception of the residues around 6 Å of dasatinib as reported in ref. 26 . To finally compute the residence times, we considered the dasatinib in the unbound state when is surrounded of a water shell having a radius of 6 Å. To assess the statistical significance of our results we performed a bootstrap analysis on the simulated unbinding times as reported in Mollica et al.²⁶

4. Distance fluctuation analysis

We analyzed the distance fluctuation (DF_{ii}) along MD simulations of four molecular systems (i.e. KD_{in} , $T315-KD_{in}$, Myr/KD_{in} and $Myr/T315-KD_{in}$) as reported in Morra et al.²⁸ in order to obtain useful insights into their allosteric behavior. In brief, DF_{ii} is defined as:

$$
DF_{ij} = \langle (r_{ij} - \langle r_{ij} \rangle)^2 \rangle
$$

where r_{ij} corresponds to the distance between C α atoms or residues *i* and *j*. Low DF_{ij} values reflect a more efficient residues communication during the trajectory.

Supporting Figure 1. Representation of atoms facing the three orthosteric pockets in PNP. The first and the second orthosteric pockets are represented by pID 9 (blue spheres) and pID 12 (green spheres). The union of pID4 and pID 13 (red and yellow, respectively) represent the third orthosteric pocket.

Supporting Figure 2. Pockets network on the PNP system. Each pocket (i.e. node of the network) is represented as a sphere, while the black line (i.e. edge of the network) represents the communication between pockets. Only pockets with a persistency higher than 30% of the simulation time are represented.

Supporting Figure 3. Left: Smoothed volume of pockets within the ATP binding site of KD_{in} (upper panel) and KD_{out} (bottom panel). Right: Representation of the atoms that enclose the pockets reported in the left plot.

Supporting Figure 4. Solvent Excluded Surface (SES) based pocket definition. The pocket definition here follows the Richards-Lee definition of the solvent excluded surface, that is the surface obtained by rolling a probe over the atomic system as shown in the left panel. [ref Richards] In our case, the pockets are identified by performing the volumetric difference between the regions enclosed by the SESs obtained with a larger probe, (right panel, in orange), and a smaller one. In our case the radii values are 3\AA and 1.4\AA , respectively. The resulting regions are then duly polished, in order to eliminate spurious ones.

Supporting Figure 5. Distance fluctuation matrix of KD_{in} , T315-KD_{in}, Myr/KD_{in} and Myr/T315- KD_{in} systems.²⁸ The red boxes in each matrix highlight the distance fluctuation between the ATP (cyan sequence) and the myristate (violet sequence) pockets. The color code represents the intensity of the distance fluctuation. White spots correspond to low fluctuation and, therefore, efficient communication between two protein regions; blue spots rather indicate a high fluctuation and a poor communication between two protein regions. As expected, distance fluctuation analysis indeed confirms a communication between the ATP and the myristate pockets in KD_{in} , Myr/KD_{in} and in Myr/T315I-KD_{in} systems. In T315I-KD_{in}, this analysis also confirmed that the communication between the ATP and the myristate pockets is weakened, in line with our results.

Supporting Table 1. List of residues and associated frequency for pID 5, pID 28 and pID 3 of KD_{out} system and for pID 5 and pID 3 of KD_{in} system.

Replicates	Das/T315I- KD_{in} Time [ns]	Das-Myr/T315I- KD_{in} Time [ns]
$\overline{2}$	19.7	5.9
$\overline{3}$	27.9	14.8
$\overline{4}$	45.9	45.8
5	9.8	39.3
6	40.9	5.6
$\overline{7}$	17.7	66.9
8	19.7	112.2
9	4.8	14.3
10	17.5	43.8
11	36.0	21.1
12	6.9	13.6
13	3.2	24.9
14	19.1	17.4
15	26.9	23.6
16	14.0	13.1
17	9.1	20.9
18	37.3	55.8
19	23.2	12.9
20	17.9	28.0
Average	17.3 ± 2.8	30.6 ± 5.6

Supporting Table 2. Residence time of dasatinib expressed in ns for Das/T315I-KD_{in} and Das-Myr/T315I-KDin systems. The average values and the bootstrap standard deviation are reported on the bottom of the table result from bootstrap analysis.

REFERENCES

- (1) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. AMBER 10. University of California: San Francisco 2008, p AMBER 10, University of California, San Francisco.
- (2) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03. Gaussian, Inc: Wallingford CT 2004, p Gaussian 03, Gaussian, Inc., Wallingford CT.
- (3) Hirschi, J. S.; Arora, K.; Brooks, C. L.; Schramm, V. L. Conformational Dynamics in Human Purine Nucleoside Phosphorylase with Reactants and Transition-State Analogues † . *J. Phys. Chem. B* **2010**, *114* (49), 16263–16272.
- (4) Shi, W.; Ting, L. M.; Kicska, G. A.; Lewandowicz, A.; Tyler, P. C.; Evans, G. B.; Furneaux, R. H.; Kim, K.; Almo, S. C.; Schramm, V. L. Plasmodium Falciparum Purine Nucleoside Phosphorylase: Crystal Structures, Immucillin Inhibitors, and Dual Catalytic Function. *J. Biol. Chem.* **2004**, *279* (18), 18103–18106.
- (5) 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.* **2004**, *25* (9), 1157–1174.
- (6) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247– 260.
- (7) Decherchi, S.; Berteotti, A.; Bottegoni, G.; Rocchia, W.; Cavalli, A. The Ligand Binding Mechanism to Purine Nucleoside Phosphorylase Elucidated via Molecular Dynamics and Machine Learning. *Nat. Commun.* **2015**, *6*, 6155.
- (8) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (8), 1950–1958.
- (9) 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.* **1983**, *79* (2), 926–935.
- (10) Congreve, M.; Andrews, S. P.; Doré, A. S.; Hollenstein, K.; Hurrell, E.; Langmead, C. J.; Mason, J. S.; Ng, I. W.; Tehan, B.; Zhukov, A.; Weir, M.; Marshall, F. H. Discovery of 1,2,4-Triazine Derivatives as Adenosine A $_{2A}$ Antagonists Using Structure Based Drug Design. *J. Med. Chem.* **2012**, *55* (5), 1898–1903.
- (11) Liu, W.; Chun, E.; Thompson, A. A.; Chubukov, P.; Xu, F.; Katritch, V.; Han, G. W.; Roth, C. B.; Heitman, L. H.; IJzerman, A. P.; Cherezov, V.; Stevens, R. C. Structural Basis for Allosteric Regulation of GPCRs by Sodium Ions. *Science (80-.).* **2012**, *337* (6091), 232–236.
- (12) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865.
- (13) Nagar, B.; Hantschel, O.; Young, M. A.; Scheffzek, K.; Veach, D.; Bornmann, W.; Clarkson, B.; Superti-Furga, G.; Kuriyan, J. Structural Basis for the Autoinhibition of c-Abl Tyrosine Kinase. *Cell* **2003**, *112* (6), 859–871.
- (14) La Sala, G.; Riccardi, L.; Gaspari, R.; Cavalli, A.; Hantschel, O.; De Vivo, M. HRD Motif as the Central Hub of the Signaling Network for Activation Loop Autophosphorylation in Abl Kinase. *J. Chem. Theory Comput.* **2016**, *12* (11), 5563–5574.
- (15) Young, M. A.; Shah, N. P.; Chao, L. H.; Seeliger, M.; Milanov, Z. V; Biggs, W. H.; Treiber, D. K.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J.; Sawyers, C. L.; Kuriyan, J. Structure of the Kinase Domain of an Imatinib-Resistant Abl Mutant in Complex with the Aurora Kinase

Inhibitor VX-680. *Cancer Res.* **2006**, *66* (2), 1007–1014.

- (16) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins Struct. Funct. Bioinforma.* **2006**, *65* (3), 712–725.
- (17) Harvey, M. J.; Giupponi, G.; Fabritiis, G. De. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, *5* (6), 1632– 1639.
- (18) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. . Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.
- (19) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593.
- (20) Vanden-Eijnden, E.; Ciccotti, G. Second-Order Integrators for Langevin Equations with Holonomic Constraints. *Chem. Phys. Lett.* **2006**, *429* (1–3), 310–316.
- (21) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 14101.
- (22) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.
- (23) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613.
- (24) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472.
- (25) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190.
- (26) Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of Protein-Ligand Unbinding via Smoothed Potential Molecular Dynamics Simulations. *Sci. Rep.* **2015**, *5* (1), 11539.
- (27) Mollica, L.; Theret, I.; Antoine, M.; Perron-Sierra, F.; Charton, Y.; Fourquez, J.-M.; Wierzbicki, M.; Boutin, J. A.; Ferry, G.; Decherchi, S.; Bottegoni, G.; Ducrot, P.; Cavalli, A. Molecular Dynamics Simulations and Kinetic Measurements to Estimate and Predict Protein–Ligand Residence Times. *J. Med. Chem.* **2016**, *59* (15), 7167–7176.
- (28) Morra, G.; Verkhivker, G.; Colombo, G.; Gunsteren, Wf.; Mark, A. Modeling Signal Propagation Mechanisms and Ligand-Based Conformational Dynamics of the Hsp90 Molecular Chaperone Full-Length Dimer. *PLoS Comput. Biol.* **2009**, *5* (3), e1000323.