Molecular modeling of complexes of PrkC protein kinase with possible endogenous targets of phosphorylation activity.

Marcin Augustyniak

2017

Abstract

'Wet' (*in vivo* or *in vitro*) experiments report CpgA (YolQ), a circularly permuted GTPase, as a potential endogenous substrate for the serine-threonine protein kinase PrkC from *B. subtilis* [1, 4, 13]. PrkC is a trans-membrane protein implicated in sporulation and biofilm formation, with cytoplasmic catalytic domain and extracellular sensor domain, and its structure and function homologous to TGF-β receptor [10, 11].

My initial results of modeling PrkCc $\cdot$ Mg^{2+} $\cdot$ ATP $\cdot$ CpgA complexes, although not completely contradictory to reports of CpgA's Thr166 [13], Thr192 or Ser226 [1] as phosphorylation sites, have significantly indicated possibility that other CpgA's residues - most prominently Thr205 - may be targets of PrkC's enzymatic activity, and have triggered obvious curiosity why and how computational results might not match data obtained from 'wet' experiments, so distantly. Surprisingly, early results has indicated also several Tyr residues - especially Tyr240 and C-terminal Tyr298 - as sites of phosphorylation, also by PrkC.

Classical molecular dynamics (cMD) simulations of PrkCc $\cdot$ ATP $\cdot$ CpgA or PrkCc $\cdot$ Mg^{2+} $\cdot$ ATP $\cdot$ CpgA complexes targeted to Thr205, have given insight into the dynamics of the two protomers, but also have questioned feasibility of obtaining reliable models of PrkCc $\cdot$ Mg^{2+} $\cdot$ ATP $\cdot$ CpgA complexes with the early approach, i.e. by rigid-body docking with subsequent cMD of selected (best) structures. The arrangement of the substrate Thr205's hydroxyl oxygen, ATP's γ-phosphorus, and the key PrkC's residues has not improved over the course of the simulations, 5 ns long each. However, the results of the simulations are in agreement with widely reported indispensability of a Mg^{2+} cation for protein kinases' enzymatic activity, as the simulation without the Mg^{2+} cation complexed with ATP has revealed substantial disruption of spatial relation within PrkC's catalytic cleft. I have explored applicability of LMOD [6–8] flexible protein-protein docking to the system under study but given its (the system) size and computational cost of LMOD flexible docking procedure I have continued pursuit for more efficient methods, reserving LMOD flexible docking, as well as cMD for the final stage structure refinement.

Majority of current protein-protein docking programs and/or algorithms bear resemblance to, if not try to mimic the 'induced fit' paradigm [9]. The energy landscape theory

[5, 15], and the conformational selection and population shift model of protein-protein recognition and interaction postulate that interacting protein partners recruit from already present in the solution populations, not necessarily the lowest energy ones, with subsequent re-equilibration (shift) of the populations [2].

Application of accelerated molecular dynamics (aMD) to conformational space sampling makes it feasible to sample conformational space of a protein nearly exhaustively starting from its near-native, e.g. crystallographic, low-energy conformation [12], and advancement in GPU hardware design has made such computations available on a workstation equipped with a modern GPU card [14].

In this work I propose a protein-protein complex structure prediction and modeling protocol that try to exploit the conformational selection and population shift model of protein-protein recognition and interaction [2] by (i) sampling conformational space of protomers separately, (ii) determination of distinct conformations and their representative structures, (iii) combinatorial docking of all/selected combinations of protomers' conformations, (iv) and refinement of selected complex structures.

I have applied this approach to PrkCc and CpgA by sampling their conformational spaces separately with aMD simulations utilizing Amber16's [3] pmemd implementation for GPUs (pmemd.cuda), and subsequent trajectory clustering in order to determine distinct conformations. I have rigidly-docked the resulting conformations and explored the vast amount of data in the search of structures with possibly optimal arrangement of the substrate, $Mg^{2+} \cdot ATP$, and the key residues of the enzyme. As the fundamental criterion of complex structure selection I have assumed geometrical constraints dictated by the typical for majority of eukaryotic protein kinases mechanism of phosphoryl group transfer, i.e. dissociative, direct, in-line, with inversion of chirality at the phosphorus atom of the transferred phosphoryl moiety.