This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard **Terms & Conditions** and the **Ethical guidelines** still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Structural analysis of MARK4 and exploration of specific inhibitors for MARK family: A computational approach to obstruct the role of MARK4 in Prostate Cancer progression

Pranitha Jenardhanan[1], Jayakanthan Mannu[1], Premendu P. Mathur[1,2]*

[1]Centre for Bioinformatics, Pondicherry University, Puducherry India
[2]KIIT University, Bhubaneshwar, Odisha, India

Corresponding Author Information

Prof. P. P. Mathur
Vice – Chancellor
KIIT University
Bhubaneshwar - 751024
Odisha, INDIA
E-mail: ppmathur@kiit.ac.in, ppmathur@gmail.com

Molecular BioSystems Accepted Manuscript

Table of Contents

Binding of identified ATP competitors specific to MARK4, characterized by unique DFG Asp-in/αC helix-out inactive state, hampers progression of prostate cancer.
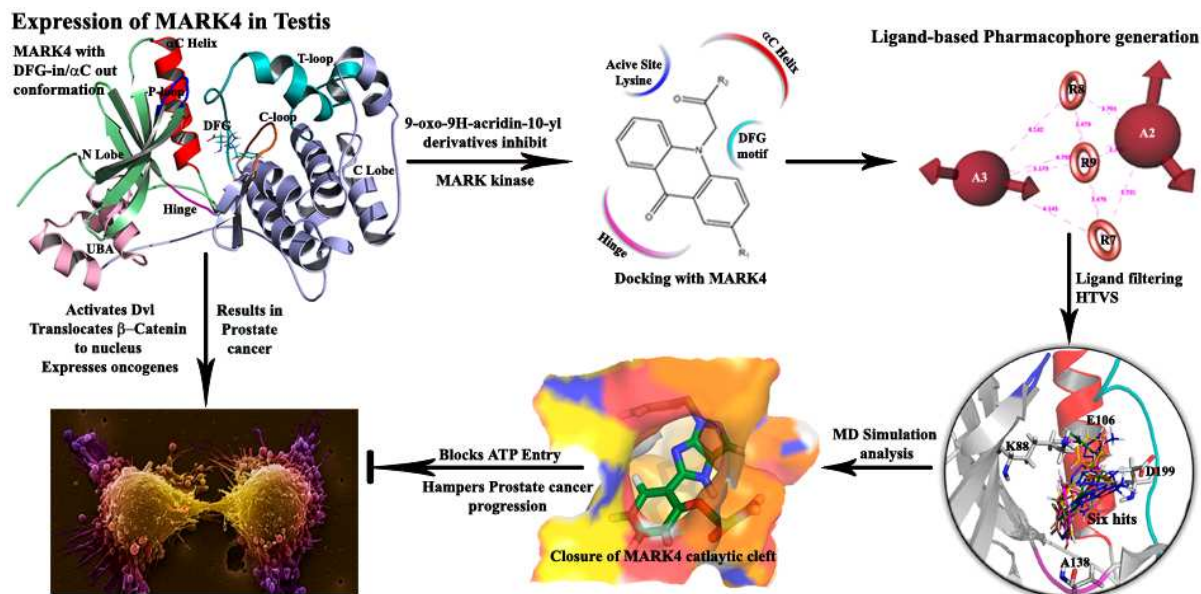
Abstract

Prostate cancer, the second most common form of cancer in adult males is generally treated using hormone therapy but emergence of hormone refractory prostate cancer poses serious challenges to the existing therapeutic strategies. However, protein kinases are now currently identified as potent targets for treating cancer, and MARK4L, a Ser/Thr kinase in the Par-1 family, is one such kinase that is expressed primarily in testis and is involved in regulation of spermatid polarity during spermatogenesis. It is also associated with Wnt-induced prostate carcinogenesis, making it a promising target for the development of anti-cancer therapeutics as part of alternative therapy to counter prostate cancer. In the present work, we predicted three dimensional structure for the kinase domain of MARK4 and analyzed its structural properties. The results illuminate the presence of unusual DFG Asp-in/αC helix-out conformation along with absence of additional hydrophobic pocket adjacent to the ATP binding site in its inactive state. These structural features accentuate the need for new specific therapeutics against MARK4. Hence, a robust ligand-based pharmacophore model AARRR.9 was developed based on the three dimensional chemical features of 9-oxo-9H-acridin-10-yl derivatives which possess high specificity towards MARK kinases. Pharmacophore based search identified six potent compounds with better specificity and binding efficiency to MARK4 bearing stable interactions with key residues K88, A138, D199 and E106, thereby making them tough ATP competitors. The closure of catalytic cleft observed in the ligand bound complexes and its independency to the movement of T-loop makes them promising candidates in hampering the role of MARK4 in prostate cancer.

Keywords: MARK, MARK4, Prostate cancer, kinase inhibitors, ATP competitive inhibitors

Introduction

Prostate cancer form of adenocarcinoma is a common neoplasm found in male prostate glands that accounts for the second most causative for death in men. Malignancy of prostate cancer increases with aberrant Wnt signaling which is involved in regulating the prostate morphogenesis, luminal epithelial cell differentiation and proliferation of prostate epithelial progenitor cells[1]. In prostate cancer cells, several gain of function mutations were observed in β-catenin and are linked with increase in the expression of its target genes through β-catenin/TCF-dependent transcription activity[2]. Of various target genes, decreased expression of MARK4L, an isoform of MAP/microtubule affinity-regulating kinases (MARK) is associated with the down regulation of β-catenin/TCF-dependent transcription activity. The expression of MARK4 is also correlated with nuclear accumulation of β-catenin in hepatocellular carcinogenesis[3].

MARK4 belongs to the family of serine/threonine kinases and is the mammalian homolog of (Partition-defective 1) PAR-1, which is involved in regulating various cellular events[4]. MAP/microtubule affinity-regulating kinases (MARK) share homology with AMP-activated protein kinases (AMPK) and have four isoforms (MARK1-4)[5-7]. MARK proteins possess an N-terminal header domain followed by the kinase domain, an ubiquitin associated (UBA) domain, a spacer region which is differentiated and natively unfolded in nature, and a C-terminal region which has a Kinase Associated (KA1) domain[8-10]. They were first identified for their ability to phosphorylate Microtubule Associated Proteins (MAPs) and are found to phosphorylate Tau and related MAPs on their microtubule binding domain, which disrupts their attachment with microtubules and disturbs the microtubule dynamics. Hyper-phosphorylation of Tau on Ser262 by these proteins is observed to correlate with increase in the progression of Alzheimer's disease[5]. Unlike other MARK isoforms, MARK4 is unique in its ability to exhibit direct association with microtubules[7]. MARK4 has two splice variants, a long isoform MARK4L with 752aa and a short isoform MARK4S with 688aa. MARK4S is expressed in neurons, while MARK4L is predominantly expressed in testis, where it is expressed in both germ cells and sertoli cells. Its expression is linked to maintain the structural changes of both apical and basal ectoplasmic specialization (ES) which is essential for the polarity of the elongating/elongated spermatids and their attachment with the sertoli cells, through its interaction with tubulin-cytoskeleton network[11].

MARK4 also plays a role in Wnt signaling and down-regulation of MARK4L, associated with the decrease in β-catenin/TCF-dependent transcription activity, identifies it as a WNT target gene. Its homology with PAR-1, a molecular switch regulating the cytoplasmic level of β-catenin, accentuates the fact that MARK4L can also act as a regulator of Wnt signaling pathway. MARK4L is up-regulated in glioma[12], hepatocellular carcinomas[3] and metastatic breast carcinomas, and it is also found to be a negative regulator of mTORC1, all of which classifies MARK4L as a potent target for treating cancer[13,]

4

[14]. The predominant expression of MARK4L in testis and its role in up-regulation of Wnt signaling pathway in prostate cancer emphasizes the fact that inhibition of MARK4L could actively hamper the progression of prostate cancer. With the changeover from hormone-sensitive to hormone-refractory form of prostate cancer being on the rise[15], the need for design of new therapeutic strategies for effective treatment has become an absolute necessity. Thus, the strategy of targeting kinases involved in signal transduction offers possible alternatives for treating cancer and many studies prove it is a therapeutically powerful approach. In the present scenario, identifying new therapeutics with better efficiency to obstruct the role of MARK4 will help to counteract prostate cancer.

Accordingly, in this study, we predicted the structure for kinase and UBA domain of MARK4L and molecular dynamics simulation was used to study its structural features and stability. Ligand based pharmacophore approach, in conjunction with virtual screening was used to identify best interacting compounds specific to MARK4 for development of potential anti-cancer drug candidates. These lead compounds were further tested for their ability to reside in and to induce changes in MARK4 binding pocket using molecular dynamics simulation.

Materials and methods

Prediction of three dimensional structure of MARK4

The sequence of human MARK4 long isoform (L) was retrieved from UniProt (ID: Q96L34)[16] and the region for Kinase-UBA domain (322 amino acids) was subjected to BLASTp[17] search against Protein Data Bank[18] database. Based on the high sequential conservation observed among the members of MARK family (Fig. 1A), we selected crystal structure of MARK3 kinase domain (PDB ID: 3FE3) (*Crystal structure of the kinase MARK3/Par-1: T211A-S215A double mutant. Nugoor C. et al., 2011. To be published.*) as template and the tertiary structure prediction was done using Discovery Studio v3.1 (Accelrys Software Inc., Discovery Studio Modeling Environment, Release 3.1.0.11157, San Diego: Accelrys Software Inc., 2005-2011.). The query sequence was aligned to template sequence by "Align Sequence to Templates" module and the obtained sequence alignment was then used to predict the structure of MARK4 using "Build Homology Models" module which uses embedded MODELLER[19]. A default of five models was generated and the best model was selected with a DOPE score of -77578.023438 Kcal/mol. Any segment of protein model in the disallowed region of Ramachandran Plot was refined using "Loop Refinement" module which uses the MODELLER energy function. The validity of the 3D structure was assessed using Ramachandran Plot in PROCHECK[20], which showed that about 93.4% of residues were in the favored region, with 6.6% of residues in the additionally allowed region, and the compatibility of 3D structure of the model with the sequence of residues it contains (1D-3D) was

assessed using Verify3D[21] and results indicate that the predicted model is well correlated with the structural features defined in its sequence (Suppl. Fig. S1).

Molecular docking studies between MARK4 and ATP and, MARK 2&4 with known 9-oxo-9H-acridin-10-yl derivatives

The refined optimized conformation of MARK4 was used for analyzing its interactions with ATP as well as with the experimentally proven inhibitors. Molecular docking was carried using Glide module of Schrödinger suite[22]. Initially the protein was subjected to Protein Preparation Wizard, where the bond orders were assigned, hydrogens added, missing loops and side chains were fixed. The prepared protein was typed with OPLS_2005 force field and energy minimized. The receptor grid for docking was generated around the catalytically key residues K88 and D181 of MARK4. The three dimensional structure of ATP and known reference inhibitors were generated using LigPrep v2.5 (LigPrep, version 2.5, Schrödinger, LLC, New York, NY, 2011.). Initially, in case of reference compounds, their two dimensional structures were drawn in ChemBioDraw Ultra v12.0 (ChemBioOffice, 2009, www.cambridgesoft.com) and then imported and prepared using the LigPrep module. The ligands were typed with OPLS force field and their possible ionization states were generated at the target pH of 7.0 with the inclusion of options for generating the isomers and tautomers. At most 32 possible combinations were generated for each ligand with at least 1 low energy ring conformation per ligand. Later, the prepared ligands were docked into the defined grid using the Glide receptor docking method, where the docking was carried out in an extra precision (XP) mode. The resultant conformations were scored using Glide XP score and further subjected to Prime MM-GBSA (Prime, version 3.0, Schrödinger, LLC, New York, NY, 2011.), in which the approximate ligand binding free energy of complexes was evaluated. Further to validate the poses obtained with MARK4, these experimentally proven compounds were also docked with crystal structure of MARK2 [PDB ID: 2WZJ][10] which was prepared as per as the protocol defined earlier and receptor grid was defined around the residues K82, D175.

Pharmacophore Model Development and Hypothesis

The 9-oxo-9H-acridin-10-yl derivatives with very low inhibition constant values (IC_{50}) were used to develop a robust pharmacophore model. The three dimensional structure of the ligands were optimized using LigPrep v2.5 and the development of pharmacophore model and hypothesis testing was carried out using Phase v3.3 (Schrödinger, LLC, New York, NY, 2011). Their biological activity, IC_{50} was converted to –log [concentration] and a "Pharma set" was created with these ligands and indicated if a ligand is set as active or inactive with respect to building the pharmacophore, and if a required match with that ligand

is essential while searching for common pharmacophores. For the development of common pharmacorphores, activity threshold of -0.5 (Active: >-0.5; Inactive: <-0.5) was set based on their IC_{50} values. In the next step, the built-in set of six pharmacophore features in Phase (Hydrogen bond acceptor – A, Hydrogen bond donor – D, Hydrophobic group – H, negatively charged group – N, positively charged group – P, Aromatic ring – R) was used to create 'site points' for each conformer of each ligand. A search was performed to find common pharmacophores among the active ligands using these site points as features and pharmacophore variants were generated, which will become a pharmacophore hypotheses, explaining how ligand binds to receptor. Derivatives 30019, 30195, 30199 were set as active and given as a criteria for required match and 30197 was set as inactive and sites were created for the given set pharmacophore features. Among the common pharmacophore models, the variants were defined to have a maximum of 5 site points and a minimum of 4, and the condition was set that at least 3 of 3 active groups must be matched. The desired variant was selected based on the number of site points included in the defined hypothesis covering more pharmacophore features. The hypotheses developed were scored using a set of very similar pharmacophores using the active set ligands. The best aligned ligand and its pharmacophore was taken as reference in order to assign a score and other actives were aligned to the reference pharmacophore using standard least-squares procedure. The scoring was measured using Alignment score, Vector score and Volume score as described in Phase protocol. If any other non-reference ligand scored better than the reference, that ligand and its pharmacophore yielding the best alignment was taken as the new reference. When all pharmacophores were scored this way, best hypothesis for each ligand-pharmacophore with the best multi-ligand alignment was taken up for the scoring process. The final survival score for each hypothesis was calculated using the scores and weights for volume, vector and relative energy parameters defined in Phase protocol. The hypotheses were filtered based on Alignment score, Site Measurements, Feature Matching Tolerances and the hypothesis with the best match to reference ligand and highest survival score was taken up to search and identify new therapeutics against MARK4.

Virtual Screening

The best scored and validated hypothesis was used to search new compounds having the structural characteristics defined in the pharmacophore model. In order to obtain inhibitors with good specificity towards MARK4, small molecules from specialized kinase focused screening libraries, Kinacore and Kinaset from ChemBridge database (http://www.chembridge.com), were used to find the best matches using "Find Matches to Hypothesis" application in Schrödinger Phase v3.3 against the pharmacophore model developed. The resulting hits were refined based on their ability to match all the site points in the

hypothesis and were selected based on their Align score and Fitness score with a matching score not less than 2.0Å distance matching tolerance, and with vector score and volume score not less than -1.0 and 0.0. The resulting compounds were taken up for Virtual Screening Workflow (VSW) in Schrödinger, which includes ligand optimization using LigPrep v2.5, ADME filtering by QikProp (QikProp, version 3.4, Schrödinger, LLC, New York, NY, 2011.), filtering against reactive functional groups followed by docking at three accuracy levels, HTVS (High Throughput Virtual Screening), SP (Simple Precision) and XP (Extra Precision). For the docking step, the filtered ligands were docked into the receptor grid of simulated MARK4 kinase domain around the residues mentioned earlier. The docking protocol was set default with post-docking minimization, and at each accuracy level, 50% of the best scoring states were taken for the next accuracy level. After the final XP docking, ligand binding energies and ligand strain energies of the complexes were calculated using Prime v3.0 MM-GBSA.

Molecular dynamics simulations of inactive ligand-free MARK4

The stability of the predicted MARK4 Kinase-UBA domain was analyzed using GROMACS 4.5.5 molecular dynamics suite[23]. Initially the structure was typed with AMBER99 protein force field[24], and the protein was placed in a triclinic box and solvated using SPC216 water model. Charges were neutralized and then the model was subjected to 1000 cycles of energy minimization using steepest descendent algorithm, followed by 1000 cycles of minimization using conjugate gradient algorithm. The movements in protein were restrained using the position restrain method and the system was equilibrated for 1ns time period in both NVT [standard canonical ensemble] and NPT [isothermal-isobaric ensemble] conditions. V-rescale method with reference temperature of 310K was employed for temperature coupling and Parrinello-Rahman method with time constant of 2.0ps and reference pressure of 1.0 bar was used for pressure coupling. LINear Constraint Solver [LINCS] constrain algorithm method with the LINCS order of 4 was used to constrain the position of bonds and Grid method with rlist and rcoulomb of 1.0nm and rvdw of 1.4nm was used. Later, the final production molecular dynamics was carried for 10ns, with the inclusion of aforementioned parameters. The obtained trajectory was then analyzed for the presence of dominant motions using Principal Component Analysis (PCA), where the g_covar option was used. Besides this, the backbone Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) were also calculated using the inbuilt options of Gromacs. Finally, with the backbone RMSD cut of 0.1nm, g-cluster option was used to cluster the obtained structures and the top cluster with least RMSD was selected for further analysis. In addition, the structures retrieved at the time scale of 500ps from the total 10ns trajectory was used for analyzing the structural transitions during MD simulation.

Molecular Dynamics Simulation studies of MARK4-identified hits

In order to evaluate the ability of the identified hits to remain in the MARK4 catalytic cleft and to study their conformational changes, we performed a 10ns molecular dynamics simulation on the complex of MARK4 bound with finally selected six compounds using GROMACS v4.5.5 suite. Initially, the ligand topologies for these compounds were prepared using PRODRG server[25], where the three dimensional structure of the bound conformation of ligands were given as inputs and they were typed with GROMOS 53A6 force field. This was followed by topology generation of proteins using the inbuilt options in Gromacs. Protein was typed with same GROMOS 43A1 force field and placed in the triclinic box and was subjected to energy minimization, position restrained equilibration runs and final production runs with the same parameters employed for simulation of apo form of MARK4. Finally, backbone RMSD graph of protein and ligand RMSD plots were generated to understand the stability of the protein–ligand complexes and the stability of hydrogen bond interactions was analyzed using the g_bond option in Gromacs.

Results and Discussion

Three Dimensional Structure of MARK4 Kinase-UBA Domain

The crystallographic structures of MARK members determined so far have provided sufficient details about the Kinase-UBA domain of MARK1-3 proteins, whereas the structure of MARK4 is still unrevealed. Considering its predominant expression in testis and its significant role in Wnt signaling, we have predicted the three dimensional structure of MARK4L and the present model represents the kinase domain in its inactive state (Fig. 1B). Superimposition with other isoforms showed that the predicted model retained all the essential features of MARK family [with MARK2, RMSD: 0.751Å; with MARK3, RMSD: 0.117Å] (Fig. 1C). The final validated model was subjected to molecular dynamics simulation for 10ns and the backbone RMSD plot with values bound between 0.24 to 0.22nm shows that the structure is stabilized after the 6thns (Fig. 2B). Superimposition of conformation retrieved from top cluster with the predicted model showed very less structural deviations with an RMSD of 1.287Å (Fig. 2A) and projection of 1, 5, 10 and 20th Eigen vectors also denotes that the predicted model has attained its stability in the given timescale (Fig. 2D). RMSF plot of the residues indicates that the residues of T Loop and αC helix are the most fluctuating regions, ranging from 0.2 to 0.29nm (Fig. 2C), which corroborates with the inherent dynamic quality of these regions. The conformation at the 10thns was selected to assess its quality over the predicted model, and results of both PROCHECK and Verify3D substantiate that the quality is further enhanced in simulated structure (Suppl. Fig. S1).

Structural features of MARK4 Kinase-UBA domain

The structure for inactive state of MARK4L [P50-E372] presented here comprises of a fragment of N terminal header [P50-N58], a catalytic domain [Y59-I310], a linker region [N311-E318], a common docking motif (CD) [E314-E319] and UBA domain [Y322-E372] (Fig. 1B). The catalytic domain holds a highly conserved structural architecture as seen in other Ser/Thr kinases; bi-lobed structure with a smaller N-lobe and a larger C-lobe and catalytic cleft between them. N-Lobe [Y59-V134] is dominated by β sheets, with five β-strands and one helix αC, while C-lobe [E142-I310] is made up of six α-helices. These lobes are connected by a flexible hinge region [M135-G141] which dictates the opening and closing of the catalytic cleft. The catalytic cleft is wide open in inactive MARK4 with an area of 1086.9Å^2 and a volume of 1157.6Å^3. As in related kinases, the strands β1-3 and helix αC of N-Lobe, hinge region and activation loop of C-lobe surrounds this cavity.

The essential functional units of catalytic domain include P-Loop, hinge region, activation loop (T-Loop), and a catalytic loop (C-Loop), which are crucial for binding of ATP into the cleft[26]. The loop connecting strands β1-β2 is termed as the phosphate binding loop (P-Loop) [K67-A71] and forms the roof of catalytic cleft. It is flexible and in active form, it helps in positioning the γ-Phosphate of ATP for transfer to substrate[9]. The RMSF plot obtained from simulation studies confirms the flexibility of the residues in this region. Mutation of catalytic residue, K88 to alanine, annuls the kinase activity and so is postulated to be the active site residue of MARK4[27] which resides in the β3strand of N-lobe. Hinge region holds approximately five residues, of which there are two conserved glycine residues in other MARK proteins, while in MARK4 the first glycine is replaced by alanine (A140). The loop extending between the residues D199-E225 in the C-lobe is termed as the activation-loop or T-loop and plays a crucial role in activating MARK4 through phosphorylation of T214[7]. T-Loop has the characteristic DFG motif[28] (Asp-Phe-Gly) in which conserved aspartic acid [D199 in MARK4] is essential in coordinating ATP through direct interaction and/or through coordinating the magnesium ion with ATP. This loop adopts an open conformation in the active kinases, while in the inactive state it adopts varied conformations and is untraceable by crystallographic methods. In the predicted inactive MARK4 kinase structure, residues N204-D213 of T-Loop are folded into the cleft and this conformation is stabilized through the interactions between the residues F206, L208, E205 in T loop with A71, K67, I70 of β3 strand of N-lobe. The loop connecting strand β7 to helix-E is termed as the catalytic loop (C-loop) [K175-L187], and the presence of RD motif [R180-D181] groups MARK4 as a member of RD kinases. On activation, this conserved arginine is observed to interact with the primary phosphorylation site in T-Loop[29]. In the predicted MARK4 model, the primary phosphorylation site T214 and R180 of RD motif are placed at a distance of 9.5Å confirming the inactive state of MARK4.

UBA domain: Ubiquitin Associated Domain (UBA) is a small domain with three helices α1-3 (Y322-E372). Helices α1 and α3 are placed antiparallel to each other forming the characteristic U shaped[30]. UBA establishes a strong hold over N-Lobe which aid in its auto regulatory function and helps to maintain the catalytic domain in inactive state through strong hydrophobic interactions mediated by the residues of helix α3 in UBA domain with β2, β3 and β4 of N-lobe. The region of MARK4 connecting the UBA domain with catalytic domain holds the conserved common docking motif (CD) [E314-E319] and a linker region [N311-E318].

MARK4 with unique inactive state conformation

In general, inactive state of kinases is universally marked by structural signatures which involve the folding of T-Loop into the cleft as well as orientation of DFG motif (Asp-Phe-Gly). This is characterized by the presence of conserved aspartate of DFG motif facing away from the cleft with conserved phenylalanine towards the cleft, forming the well-known DFG-out mode (Fig. 4A). It is also marked by the positioning of core helix αC in the 'out' mode, evidenced by the absence of salt bridge between catalytic lysine in β3 strand and conserved glutamate in helix αC (Fig. 3). However, in active kinases, the activation starts with the phosphorylation of threonine in T Loop, which relocates this loop between helix αC and helix F and brings about a flip in the orientation of conserved DFG aspartate and phenylalanine residues, such that their positions are reversed, making aspartate to face the cleft and phenylalanine residue to face away from the cleft forming DFG-in mode (Fig. 4B). Parallel to this, a shift in the position of helix αC occurs so that it is packed against the N lobe wherein its conserved glutamate forms ionic pair with catalytic Lysine residue[31]. Though it is obvious to expect MARK members to adopt these structural features mentioned, investigation of its structural features and comparison with other members of Ser/Thr kinases reveals the presence of distinctive and unique features in inactive states of MARK kinase domains (Fig. 4C-D).

In inactive MARK kinases (inclusive of our predicted structure), T loop is observed to be folded into the cleft and placed beneath the P-Loop, and helix αC [P97-G112] is observed to adopt 'out' mode, which is evident from the absence of salt bridge between the residues K88 and E106 (Fig. 3). However, quite contrasting to the existing DFG-out mode in other inactive kinases, the DFG motif in MARK kinases displays a unique conformation. Here in MARK4, the conserved Aspartate D199 is observed to face towards the cleft and the conserved Phenylalanine F200 faces away from the catalytic cleft (Fig. 4C-D), which is very much uncommon in inactive state of kinases. Instead of adopting the classic DFG-out mode, MARK kinases exhibit peculiar DFG-in mode in their inactive state (Fig. 4E). However, the orientation of conserved Glycine (G201) plays an essential role in regulating the position of conserved

11

Aspartate (D199), through the formation of a hydrogen bond[28]. In inactive state of many kinases, this glycine will be flipped away thereby disrupting the formation of hydrogen bond with Aspartate, which is essential to place the Aspartate in proper orientation for coordinating the ATP directly or for binding with the magnesium ion. Absence of this hydrogen bond in the inactive state of kinases, induces the conserved aspartate to change its orientation away from the cleft for adopting the DFG-out mode, which upon phosphorylation of T-Loop, rotates towards the cleft and forms the DFG-in mode. MARK kinases follow this conserved mechanism in their active state, but in inactive MARK kinases, even if the hydrogen bond between the aspartate and glycine is not observed, still, this aspartate does not rotate away from the cleft, indicating that all MARK kinases in their inactive state possess the peculiar intermediate DFG-in conformation, thereby conferring specificity over other kinases. Fig. 4F shows the differences in the orientation of DFG Aspartate and Glycine between active and inactive MARK. In addition, the presence of DFG-out mode in inactive kinases opens up an additional hydrophobic pocket (HP-II) adjacent to the cleft, which is readily accessed by the type II kinase inhibitors (Fig.4A). These inhibitors require the presence of DFG-out mode and the additional hydrophobic pocket for binding and inhibition of inactive kinases[31]. The absence of both DFG-out mode as well as HP-II pocket in inactive MARK kinases signifies the difficulty in using existing Type II kinase inhibitors. Hence, the presence of unusual DFG-in mode accompanied with helix αC-out mode (DFG Asp-in/αC helix-out) makes MARK kinases unique and these features necessitate the need for specific inhibitors to bind and obstruct their function.

ATP Binding in MARK4

Although the conservation of catalytic core is well established in MARK kinases, information on the mode of binding and specific residues involved in binding of ATP are unavailable as there are no solved structures complexed with ATP. Hence, we performed molecular docking to study the interactions involved in the binding of ATP into MARK4. Docking results reveal that ATP occupies the catalytic cleft and is stabilized through formation of hydrogen bonds, charged interactions and van der Waals interactions with key residues.

As in the conventional active form of kinases, ATP occupies the catalytic cleft in MARK4 and establishes similar binding mode through the formation of two hydrogen bonds by adenine ring with A138 (1.6Å and 2.1Å), while its ribose region is anchored thorough the formation of hydrogen bond with catalytic K88 (2.3Å). Their triphosphate region extends till the core helix αC and forms charged interactions with K88 and van der Waal's contacts established with the residues of hinge region [Y137], T-Loop [F202, L208], helix C [L103] thus stabilizing the MARK4-ATP complex (Fig. 5). In analogy with related kinase structures, the ATP binding site of MARK4 is delineated into four regions: adenine binding region, ribose

binding region, phosphate binding region, and hydrophobic pockets I and II (HP-I, HP-II). Adenine binding region involves the residues of hinge region with E136 acting as the gate keeper residue. Residues in strands β1-3, as well as residues V73, A86 and K88 form the ribose binding site and the residues M110, E106 of helix αC, D199 of activation loop and K88 of strand β2 form the phosphate binding site while I65, A86 and G141, L188 form hydrophobic regions I and II respectively (Fig. 5). Since in MARK4 DFG Asp-in/αC helix-out inactive state the F200 orients away from the ATP binding pocket, it does not inflict steric hindrance on the binding of ATP which will be otherwise imposed in classic DFG-out mode in inactive kinases. This observation is quite significant and suggests the fact that inactive MARK kinases can allow the binding of ATP even if T-Loop is folded into the catalytic cleft. However, helix αC-out mode may not favor the transfer of phosphate groups to the substrate suggesting that the observed T-Loop conformation and DFG conformation may favor binding of ATP but not substrate binding.

Interaction studies of known inhibitors 9-oxo-9H-acridin-10-yl derivatives

Aberrant expression of MARK4 leading to proliferation of tumor cells essentiates the design for new inhibitors to block its function in tumor cells. In connection to this, Thomas Timm *et al.*, have experimentally identified kinase inhibitors with specificity towards all MARK kinases and less reactivity against SAD-B and AMPK kinases, using recombinant MARK2 catalytic domain[32]. A total of four derivatives with a common core group, 9-oxo-9H-acridin-10-yl moiety, denoted as 30019, 30195, 30197 and 30199 (Fig. 7A), were observed to have very low inhibition constants making them efficient but their binding mode with MARK kinases were not studied. In order to develop a ligand based pharmacophore specific to MARK4, we first studied the binding efficiency of these ligands against MARK4 using molecular docking studies and to cross-validate the results, we have also docked the experimental ligands with MARK2 kinase using optimized rat MARK2 crystal structure (PDB ID: 2WZJ).

With MARK2 kinase domain: The core, acridin-9(10H)-one moiety of these derivatives occupies the adenine binding region of catalytic cleft (Fig. S2(A)) through formation of hydrogen bond and short van der Waals contacts with residues of hinge region, and they are also observed to form van der Waals as well as charge interactions with residues in hydrophobic pocket I [I59, G60 and K61] and II [G135 and L182] of catalytic left [Fig. S2 and Table 1]. The aromatic rings of acridine in 30195, 30197 and 30199 also establish Π-Π stacking interactions with Y131 (Fig. S2(B-D)), while in 30019, presence of methoxy group in R2 position is found to distort the acridine plane towards hydrophobic pocket II, resulting in the loss of Π-Π stacking with Y131 (Fig. S2(E)). This common core moiety mimics the interactions of ATP

adenosine ring, and its three membered acridine ring offers a suitable chemical environment for formation of contacts with hinge region, thereby preventing the access of ATP into catalytic cleft. These attributes make them appropriate starting material for identification of new specific ATP competitors against MARK.

With MARK4 kinase domain: The 9-oxo-9H-acridin-10-yl derivatives were docked into the ATP binding site of MARK4 and the obtained docking scores are as good as the scores resulted from docking of these derivatives with MARK2 [Table 1], and appreciably low relative binding free energy values of the complexes infer the ability of the ligand to reside into the binding pocket (Suppl. Table 1). Analysis of the results elucidate that these derivatives also occupy the ATP binding pocket (Fig. 6A) and their ability to interact with ATP binding residues were analyzed with reference our MARK4-ATP model. As observed in MARK2 bound complexes, the core acridin-9(10H)-one moiety of 30199 and 30197 occupies the adenine binding region of MARK4 catalytic cleft (Fig. 6B-C) and is anchored through hydrogen bond with A138, and by formation of van der Waals contacts with residues E136, Y137, S139 of hinge region and V73, A86 of ribose binding pocket [Table 1]. Also, this moiety forms Π - Π stacking interactions with Y137, which was not observed in ATP binding. However, the plane of acridine rings in 31095 and 30019 is tilted by 3.5Å and 1.8Å respectively from their mean position, facing towards the hydrophobic pocket II, instead of occupying the adenine binding region of catalytic cleft. As a result, the acridine ring has failed to establish hydrogen bond with A138, while it involves in Π-cat$^+$ interaction with K88 and forms van der Waals contacts with residues of hinge region, T-Loop, C Loop and with conserved D199 of DFG motif (Fig. 6D-E). In MARK4 bound structures, the R1 functional groups of 30199, 30195 and 30019 are found to traverse the triphosphate binding region of ATP and extended till the helix αC, while in 30197, the R1-cyanophenyl group is folded and restricted to the entry point of the cleft (Fig. 6A).

9-oxo-9H-acridin-10-yl derivatives as ATP competitors: From the docking results of 9-oxo-9H-acridin-10-yl derivatives with both MARK2 & MARK4, it becomes clear these derivatives show high tendency to occupy the cleft between the lobes, and form strong contacts with key residues like K88, D199 of DFG motif and E106 of helix αC. Moreover, their core moiety has helped in forming strong contacts with A138 of hinge residue and also with residues of ribose binding pockets and hydrophobic pockets I and II of catalytic cleft. All these observations suggest that these derivatives occupy the ATP binding cavity and their scaffold impersonates the interactions of adenosine moiety of ATP and the experimental results obtained from the study of Thomas Timm *et. al.*, ascertains this observation[32]. In their study, they tested for the efficacy of these derivatives to inhibit the action of MARK in presence of

14

both ATP and substrate, and results explicates that the inhibition was affected by the increase in concentration of ATP but not affected by the presence of substrate, which confirms that these derivatives act as ATP competitors. Analysis of our docking studies strongly supports the experimental observation and defines how they could effectively inhibit the function of MARK4 and act as ATP competitors. Most importantly, less cross reactivity of these derivatives towards other kinases highlights them as potent leads, which is utilized in our study for the further development of MARK specific pharmacophore.

Development of Ligand-based Pharmacophore Model

The presence of DFG Asp-in/αC helix-out catalytically inactive conformation in MARK kinases identifies them as unique kinases, and poses difficulties in using the existing Type-II kinase inhibitors for inhibiting their function. The specificity of 9-oxo-9H-acridin-10-yl derivatives to bind with and act as ATP competitive inhibitors provides novel scaffold for the development of MARK specific inhibitors. Our docking results also substantiate the abilities of these leads to target the key residues of MARK4. Since the development of ligand based pharmacophore and their utilization to screen available drug candidates is a successful and promising approach for drug design strategies, we developed a pharmacophore model based on these derivatives to identify more efficient drug candidates. The structural characteristics responsible for their specificity and their biological activity were analyzed, and a robust pharmacophore model was built with all these features. This pharmacophore model was used to identify kinase inhibitors with improved specificity towards MARK family of kinases.

The structure of four 9-oxo-9H-acridin-10-yl derivatives with biological activity, IC_{50} were taken and a pharma set was defined based on activity threshold. A total of 7 variants were generated based on the frequencies of the features and all 7 variants were used to develop the pharmacophore hypothesis. Of these, two variants, ARRRR and AARRR, representing most of the key features of active ligands, with the former having 1 hypothesis and the latter with 6 hypotheses were scored to select the best hypothesis (Table 2). After scoring actives with these hypotheses, AARRR.9 had the top survival score of 3.357 with matches to all three actives and was selected as the best hypothesis for further studies (Table 3). The hypothesis AARRR.9 has pharmacophore sites of two H-bond acceptors (A) and three aromatic rings (R) with specific inter-site distances between these features (Table 4). The spatial arrangement and the features of the pharmacophore hypothesis are depicted in Fig. 7B-C and it correlates with the structural requisites mentioned in our docking study to block MARK kinase activity.

Pharmacophore search for new therapeutics

The best hypothesis, AARRR.9 was used for identifying kinase inhibitors with better and enhanced specificity towards MARK4. This hypothesis was screened against the ligand sets based on compounds known to possess inhibitory activity against kinases. Two ligand sets with a total of 26,266 compounds were taken from Chembridge Database: Kinaset and Kinacore, and the compounds were matched against the hypothesis. A total of 919 hits that matched all the site points were then optimized and post ligand optimization yielded 1552 compounds which includes tautomers and stereoisomers if generated. The resultant entries were checked for their ADME properties and 1508 entries cleared ADME. Of these, a total of 1174 compounds which cleared both Lipinski and Reactive functional groups filters screening were taken for High Throughput Virtual Screening (HTVS) with predefined grid of MARK4, followed by rescoring of the docked complexes using Prime MM-GBSA. HTVS resulted in 161 compounds with 80 compounds on SP mode and 40 compounds in XP docking. The ligand binding free energy was calculated for all 40 compounds and finally 6 compounds with top Glide score and relatively low binding free energy values better than the reference compounds were selected for interaction analysis and molecular dynamics simulation studies. The details of these selected ligands and their fitness scores with the best pharmacophore are provided in Table 5 and the overview of the entire methodology is flowcharted in Fig. 12.

Docking and simulation analysis of obtained top hits with MARK4

Having identified the six best scoring hits based on XP docking protocol, their interactions with MARK4 kinase domain were studied in detail using molecular dynamics simulation. Their interactions profile was compared with both ATP and the reference compounds 9-oxo-9H-acridin-10-yl derivatives. The docking results are tabulated in Table 6 and their interactions with residues of catalytic cleft are displayed in Fig. 8. The changes observed in the interaction profile of compounds during simulation are tabulated in Supplementary Table 2. The selected six compounds fall into four different categories based on their scaffolds, acetamide (Entry ID: 214,781 and 529), benzamide (Entry ID: 950), benzamidazole (Entry ID: 389) and xanthen-9-one (Entry ID: 411) moieties.

Compound 411: The top hit compound 411 possess 9H-xanthen-9-one as core moiety, with a nitrophenyl group attached through oxoethoxy group and a Glide docking score of –8.453 Kcal/mol (Table 6). Presence of 9H-xanthen-9-one moiety in this compound is similar to that of 9-oxo-9H-acridin-10-yl core moiety of reference compounds and this favors a similar kind of binding mode with hinge region of catalytic cleft, with their xanthene rings occupying the adenine pocket and 2-oxoethoxy group occupying the ribose binding region of MARK4. It facilitates the formation of hydrogen bonds with residues A138

16

and E136 of hinge region and hydrophobic contacts with residues A86 and G141 of HP-I and HP-II of the catalytic cleft. Their R group, nitrophenyl moiety, forms van der Waals contacts with residues M110, V119, A198 of phosphate binding site (Fig. 8A). On MD simulation of the complex, the ligand is observed to be retained in the binding cleft and the complex is stabilized as evident from the RMSD plot, which shows that fluctuations have attained a plateau at 0.3nm (Fig. 9A). However, ligand RMSD graph and superimposition of simulated complexes retrieved at 500ps trajectory interval with the native docked complex exhibit differences in the binding mode of ligand (Fig. 9B & 9C). In post simulated complexes, the plane of the xanthene ring is shifted by 2.1Å from its mean position leading to changes in the interaction profile of the docked complex. Specifically, the hydrogen bond formed between xanthene atom O7 and A138 is replaced by a new stable hydrogen bond between xanthene-O3 and A138. Similarly, the change in the orientation of nitrophenyl group has induced the formation of new less stable hydrogen bonds with M110 and K88 (Fig.10A-B). Nevertheless, this ligand spans the entire binding cleft and makes charged and van der Waals interactions with K88, D199 and E106 and maintained its interaction with A138 and K88. These differences accompanied by movement of N lobe towards the catalytic cleft by about 1.65Å, has induced shrinkage of the catalytic cleft, which is evident from the decrease of binding site volume from 432.86Å^3 in docked complex to 346.43Å^3 in post-simulated complex (Fig. 11A & 11B). These observations signify that the compound 411 can strongly bind into the cleft and could bring about the closure of the catalytic cleft, thus preventing the access of ATP and inhibit MARK4 effectively.

Compound 214: In MARK4-214 complex, tetrahydrobenzothiophene-3-carbonitrile moiety occupies the adenine binding region of MARK4 and its orientation is stabilized through the formation of Π-Π and Π-σ interactions with Y137 and M135 residues respectively, but it does not form hydrogen bonds with A138 as seen in reference compounds. It does involve in the formation of van der Waals contacts with residues I65, A86 and G141, L188 of HP-I and HP-II respectively (Fig. 8B). Their phenyl triazol-thio-acetamide moiety occupies the ribose and phosphate binding regions of MARK4 and extends till the core helix αC by interacting with catalytic K88, E106 and D199, and has charged and van der Waals contacts with other residues of catalytic cleft with a good docking score of -8.067 Kcal/mol (Table 6). This complex attains stability after 5ns (Fig. 9A) during simulation, and ligand RMSD plot depicts variations in value and is stabilized at the range of 0.5nm (Fig. 9B). The superimposition of native docked and simulated complexes depicts that the whole compound is translated by about 2.8Å from its mean position observed in the docked complexes (Fig. 9D), and as a result, a new hydrogen bond was formed between the residue A138 and the core moiety of compound 214. This newly formed interaction was maintained, but the hydrogen

bond between D199–N1:214 was lost during simulation (Fig. 10C). Also, Π-Π and Π-σ interactions observed with Y137 and M135 were maintained till the end of the simulation and the orientation of terminal phenyl group in the space between K88 and E106 was also maintained, which indicates that this orientation can effectively hinder the formation of salt bridge between K88 and E106. Moreover, binding of this ligand induced a shift in the position of N-Lobe with an RMSD of 2.516Å and reveals the presence of considerable drift in the positions of P-Loop, strands β2-4 and in loop connecting the strand β3 – helix αC. This leads to a decrease in the volume of binding site from 374.89 $Å^3$ to 253.82 $Å^3$ (Fig. 11A, 11C) similar to that observed in MARK4-411 complex, which effectively blocks the entry of ATP into the cleft. In addition, the specific orientation of terminal phenyl group of compound 214 between the residues K88 and E106 also suggests that this ligand can also restrain the movement of core helix, thereby affecting the activation mechanism of MARK4 and hold it in the inactive state.

Compound 950: In MARK4–950 complex, the pyrazine moiety is anchored in the adenine binding pocket and as seen in the xanthene rings of compound 411, it forms hydrogen bond with A138 with a distance of 2.25Å. The benzofuran ring attached with pyrazine moiety is tilted towards the hydrophobic pocket II and is stabilized by van der Waals contacts with residues E185 and L188. The ribose binding site is left unoccupied, but the ligand forms two hydrogen bonds with K88 and D199 and extends till the gamma phosphate binding region making charged interaction with E106 and M110 and Π-cat$^+$ interaction with K88. Unlike MARK4-214 and MARK4-411 complex, this ligand spans the complete ATP binding region and makes strong contacts with hinge region, catalytic K88, with D199 of DFG motif and with E106 of core helix αC with a Glide score of –7.386 Kcal/mol (Fig. 8C) (Table 6). This complex exhibits less structural transitions and is stabilized with an RMSD of 0.2 nm, (Fig. 9A). Moreover, in post simulated complex, the compound experiences less deviation (ligand RMSD: 0.15nm) after the 3rd ns (Fig. 9B). The pyrazine moiety is tilted by 1.7Å, while the pyrrole ring is shifted horizontally by 90° (Fig. 9E) resulting in the loss of hydrogen bond with D199, while the hydrogen bonds formed with K88 and A138 remain unchanged (Fig. 10D). In addition to existing hydrogen bonds, two new bonds formed with K88 (K88:HZ3 - UNK: O3 and K88:HZ1 - UNK:N5) were maintained with slight disturbances (Fig. 10E). The DFG motif was moved by a distance of 3.1Å from its mean position resulting in the space held between T loop and C loop to increase, leading to the rise in binding site volume from 348.48 $Å^3$ to 437.66 $Å^3$ (Fig. 11A & 11D). This observation is very much contrasting to the closure effect induced by compounds 411 and 214 but strong contacts established with A138, K88, E106 and D199 indicates that this compound can act as a strong lead molecule.

Compound 389: In this compound, the benzo(*d*)imidazole group occupies the hinge region and forms hydrogen bond with A138 and van der Waals contacts with residues A86, I65, V73 and L188. The thiadiazole ring of compound 389 is placed above the ribose binding pocket and it forms two hydrogen bonds and two Π-cat+ interactions with K88, while the terminal benzene ring spans the phosphate binding region and forms charged interactions with E106, E136, T207, L208 and van der Waals contacts with F202, F206, S203 and L103 (Fig. 8D). The docked complex has a Glide score of -7.290 Kcal/mol (Table 6) and molecular dynamics simulations show that this complex maintains its stability throughout the time period without much deviation (Fig. 9A) and the ligand RMSD plot did not indicate significant variations (Fig. 9B & 9F). However, thiadiazole moiety displays variations in its orientation resulting in the hydrogen bonds formed with K88 in the docked complex to lose its stability but has managed to maintain the Π-cat$^+$ interactions with K88 throughout the simulation period. Also, the hydrogen bond formed by the benzo(*d*)imidazole group with A138 has remained unchanged throughout simulation (Fig. 10F). As observed in 411 and 214 bound MARK4 complexes, the binding of compound 389 has induced a decrease in the binding site volume of catalytic cleft from 324.48Å^3 to 267.54Å^3 (Fig. 11A & 11E). This compound exhibits strong contacts with key residues A138, K88, E106 and D199 and since the benzo(*d*)imidazole group and the thiadiazole ring are known well for their anti-cancer properties[33, 34], their presence confers advantage in using this ligand to inhibit MARK4 action.

Compound 781 and 529: Compounds 781 and 529 share a common scaffold of dihydroimidazole moiety with difference in their R group; the thiophene group in compound 781 is replaced by pyridine ring in compound 529. This alteration has reversed the binding orientation such that in MARK4-781 complex, the phenxoy acteamide group is placed at the entry point of the cleft whereas it is replaced by the pyridine ring in MARK4-529 complex (Fig. 8E & 8F) with docking scores of –7.137 Kcal/mol and -7.126 Kcal/mol respectively. Both the complexes show overall stability during MD simulations (Fig. 9A) but fluctuations were seen in the interactions of the ligand as exhibited in the ligand RMSD plot (Fig. 9B). In MARK4–781 complex, the phenoxy group is placed in the ribose binding region with its acetamide group is oriented towards the space between T loop and C loop while the core dihydroimidazole moiety and terminal thiophene group are placed in the triphosphate binding region of catalytic cleft. This binding conformation is favored by two hydrogen bonds formed by phenoxy acetamide group with E185 and charged interactions formed with K88, E106, F202 and F206 (Fig. 8E). However, in post simulated complexes, due to the shift in phenoxy acetamide group (Fig. 9G), the initial hydrogen bonds observed in docked complex were replaced by the formation of new strong hydrogen bond with A138 (Fig. 10G) and the orientation is further stabilized by the formation of new charged interactions with D199, V119, A198

19

and M135. However, in both docked and simulated complexes, the orientation of terminal thiophene group is found to be unaltered and it has consistently maintained its interactions with E106, F202 and F206. In MARK4-529 complex, pyridine moiety is oriented over the adenine binding pocket, with the core dihydroimidazole group and terminal phenxoy acteamide group placed in triphosphate binding regions of catalytic cleft. The changes seen in the binding mode of compound 529 (Fig. 9H) has caused the hydrogen bonds [A138-H:N42 and D199:OD1-H26] formed by pyridine and dihydroimidazole moiety in docked complexes (Fig. 8F) to be replaced by a new hydrogen bond between terminal acetamide group with D199 (Fig. 10H). As seen in other selected hits, shrinkage of the catalytic cleft is observed in both 781 and 529 bound MARK4 complexes (Fig. 11A). Interestingly, in MARK4-781 complex, the cleft is closed completely with the ligand being concealed inside (Fig. 11F). The huge drift in the position of P-Loop towards the T-Loop might have facilitated for decrease in size of the cleft from 350.20Å^3 to 238.04Å^3. Another unique feature is, in case of MARK4–781 complex, the bulky aromatic rings (phenoxy group) that usually bind to the hinge region, is folded into the cleft leaving the acetamide group to form hydrogen bonds with A138. This folding of the bulky phenoxy group into the cleft provides enough space for the movement of P-loop to close the cleft opening. This feature is quite interesting and suggests that this compound can even more effectively block the access of ATP into the cleft after its binding. Similarly, binding of compound 529 with MARK4 has induced constriction of binding site volume by displacing the N lobe (Fig. 11G), and this compound exhibits good interaction with residue D199 with the possibility that it could prevent the flip of D199 in DFG motif during the transition of MARK4 from inactive to active state.

All the six identified hits exhibit good interaction profile with MARK4. These compounds bind efficiently into the catalytic cleft and they mimic most of the interactions involved in ATP binding. It is further supported by their good Glide score as well as their relative ligand binding free energy values (Suppl. Table 1). The compounds establish strong interactions with hinge region, catalytic residue, K88, and with D199 of DFG motif. Also, their terminal R groups extend deep into the pocket till helix αC, which would restrict the flexibility of helix αC during activation of kinase domain. The inference of three aromatic rings in our pharmacophore model is very well established in the obtained top hits and the presence of aromatic rings have great importance in their binding to the hinge region, specifically to A138, as well as in formation of stabilising Π interactions with the residues of hinge region. The mid part of these compounds occupies the ribose binding region, stabilized by H-bond and Π interactions with K88 and D199 and the terminal R groups bind to ATP phosphate binding regions with solid interactions with E106 at helix αC. Simulation of the ligand–MARK4 complexes has revealed the prevailing ability of

these compounds to maintain in the catalytic cleft. Analysis of complexes during the 10ns time period shows that though there is loss of old interactions and the formation of new interactions, these changes are observed in their R groups. The interaction of the aromatic rings with the hinge region remain constant throughout the simulation which emphasizes the crucial role of the pharmacophore feature, RRR, in the model and the presence of aromatic rings in these hits. Also, it is evident that the binding of these compounds induces the closure of the catalytic cleft by the pulling in N lobe β sheets. This locks the ligand from moving out; thereby further reducing the possibility of ATP competing for the same pocket, which in turn blocks the enzymatic function of MARK4.

Conclusions

The MAP/microtubule affinity-regulating kinases 4 (MARK4) is one of the four isoforms identified under MARK family of serine/threonine kinases. MARK4 is expressed as two spliced forms viz., MARK4L and MARK4S. MARK4L is principally expressed in testis and is involved in maintaining the polarity of the elongating/elongated spermatids as well as their attachment to sertoli cells. It is also a chief regulator of the Wnt signaling pathway and its uncontrolled up-regulation is associated with Wnt-induced prostate carcinoma, making it a key target for the development of anti-cancer therapeutics. Inhibitors against kinases has become a promising approach to fight cancer and has inspired us to use this approach to counter hormone-refractory form of prostate cancer by identifying specific inhibitors that precisely target MARK4 kinase.

Since there are no available structural details regarding MARK4, we predicted its three dimensional structure and analyzed the stability of the predicted model by molecular dynamics simulation. The validated structure was then compared with existing kinases and their structural characteristics of MARK4 kinase domain give a unique undertone for the MARK family of kinases. The unusual DFG Asp-in/αC helix-out conformation of the catalytic cleft observed in inactive MARK4 makes it different from other Serine/Threonine kinases. The absence of the additional hydrophobic pocket adjacent to the ATP binding site in MARK4, which is usually present in other kinases, makes it unsuitable for Type II kinase inhibitors and confers added specificity to this protein. The structural specificity observed in MARK4 has been reflected in the need for identification of better and more specific therapeutics against MARK4. Since up-regulation of MARK4 favors the survival of cancer cells, we require specific inhibitors against MARK4 and in doing so, the unique structural characteristics of its kinase domain must be duly considered. The identification of 9-oxo-9H-acridin-10-yl derivatives as promising leads against MARK kinases with less cross reactivity with other kinases was utilized to develop a robust pharmacophore model in order to identify new therapeutics against MARK4 with better specificity and

binding efficiency. Before the development of this pharmacophore model, we studied the binding of ATP with MARK4 to identify the crucial residues involved in this binding. Accordingly, we studied the role of catalytic residue, K88, D199 of DFG motif and the importance of A138 in the hinge region in establishing strong interactions with ATP. Docking of the known inhibitors, 9-oxo-9H-acridin-10-yl derivatives with MARK2 as well as MARK4 revealed that these compounds occupy the same ATP binding pocket and forms similar interactions with the key residues mentioned. The characteristics of these derivatives were taken up in developing a pharmacophore model and based on scoring against the actives, hypothesis AARRR.9 was found to be the best and it included all the required characteristics of these derivatives. This AARRR.9 pharmacophore model was used to identify new therapeutics from the specialized ChemBridge databases of existing kinase inhibitors. A total of 26,266 compounds were screened against the pharmacophore and those with matches to all features of pharmacophore model were tested for their drug-likeness, ADME, toxicity and reactive functional groups. The resulting entries were taken for HTVS and finally six compounds were chosen based on all the above criteria. The six promising compounds occupied the ATP binding pocket with better binding affinity and exhibited interactions similar to that of those observed in docking of MARK4 with ATP as well as the known 9-oxo-9H-acridin-10-yl derivatives. Simulation of the six compounds in complex with MARK4 illuminates the strong ability of these compounds to retain in the cleft and elucidates their interaction profile. The screened compounds have features analogous to ATP which is also emphasized in the developed pharmacophore model. The results also reveal the ability of these compounds to close the catalytic cavity upon their binding emphasizing that they can act as competitors for ATP and hence utilized for inhibiting the function of MARK4. Another interesting feature of these screened compounds is that, their binding affinity does not depend on the movement of T-loop as they do not form significant interactions with T-loop residues. This highlights the effectiveness of these compounds to be resolute even during the transformation of MARK4 to active state, making them tough ATP competitors and efficient candidates against MARK4.

Acknowledgements

Electronic Supplementary Information

Fig. S1: Depicts the Ramachandran Plot and Verify 3D Analysis of Predicted and Simulated model of MARK4 Structure. **Fig. S2:** Binding mode analysis of known MARK2 inhibitors, 9-oxo-10-acridinyl derivatives, docked into the ATP-binding pocket of MARK2 kinase domain (PDB ID: 2WZJ_A). **Supplementary Table 1:** Relative MM-GBSA ligand binding free energies of the compounds bound to MARK4. **Supplementary Table 2:** Details the changes observed in interaction during the MD simulation of MARK4 complex with top-scoring hits.

References

1. G. Kharaishvili, D. Simkova, E. Makharoblidze, K. Trtkova, Z. Kolar and J. Bouchal, *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*, 2011, 155, 11-18.
2. X. Yu, Y. Wang, D. J. DeGraff, M. L. Wills and R. J. Matusik, *Oncogene*, 2011, 30, 1868-1879.
3. T. Kato, S. Satoh, H. Okabe, O. Kitahara, K. Ono, C. Kihara, T. Tanaka, T. Tsunoda, Y. Yamaoka, Y. Nakamura and Y. Furukawa, *Neoplasia*, 2001, 3, 4-9.
4. K. Kemphues, *Cell*, 2000, 101, 345-348.
5. G. Drewes, B. Trinczek, S. Illenberger, J. Biernat, G. Schmitt-Ulms, H. E. Meyer, E. M. Mandelkow and E. Mandelkow, *J Biol Chem*, 1995, 270, 7679-7688.
6. S. Illenberger, G. Drewes, B. Trinczek, J. Biernat, H. E. Meyer, J. B. Olmsted, E. M. Mandelkow and E. Mandelkow, *J Biol Chem*, 1996, 271, 10834-10843.
7. B. Trinczek, M. Brajenovic, A. Ebneth and G. Drewes, *J Biol Chem*, 2004, 279, 5915-5923.
8. N. J. Bright, C. Thornton and D. Carling, *Acta Physiol (Oxf)*, 2009, 196, 15-26.
9. T. Timm, A. Marx, S. Panneerselvam, E. Mandelkow and E. M. Mandelkow, *BMC Neurosci*, 2008, 9 Suppl 2, S9.
10. A. Marx, C. Nugoor, S. Panneerselvam and E. Mandelkow, *FASEB J*, 2010, 24, 1637-1648.
11. E. I. Tang, X. Xiao, D. D. Mruk, X. J. Qian, K. W. Mok, P. Jenardhanan, W. M. Lee, P. P. Mathur and C. Y. Cheng, *Spermatogenesis*, 2012, 2, 117-126.
12. I. Magnani, C. Novielli, L. Fontana, S. Tabano, D. Rovina, R. F. Moroni, D. Bauer, S. Mazzoleni, E. A. Colombo, G. Tedeschi, L. Monti, G. Porta, S. Bosari, C. Frassoni, R. Galli, L. Bello and L. Larizza, *Analytical cellular pathology*, 2011, 34, 319-338.
13. P. N. Gabrovska, R. A. Smith, T. Tiang, S. R. Weinstein, L. M. Haupt and L. R. Griffiths, *Mol Biol Rep*, 2012, 39, 3879-3892.
14. L. Li and K. L. Guan, *J Biol Chem*, 2013, 288, 703-708.
15. F. Conrad, X. Zhu, X. Zhang, R. J. Chalkley, A. L. Burlingame, J. D. Marks and B. Liu, *J Mol Med (Berl)*, 2009, 87, 507-514.
16. C. UniProt, *Nucleic Acids Res*, 2013, 41, D43-47.
17. S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res*, 1997, 25, 3389-3402.
18. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res*, 2000, 28, 235-242.
19. N. Eswar, D. Eramian, B. Webb, M. Y. Shen and A. Sali, *Methods Mol Biol*, 2008, 426, 145-159.
20. A. L. Morris, M. W. MacArthur, E. G. Hutchinson and J. M. Thornton, *Proteins*, 1992, 12, 345-364.
21. R. Luthy, J. U. Bowie and D. Eisenberg, *Nature*, 1992, 356, 83-85.
22. T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *J Med Chem*, 2004, 47, 1750-1759.

23.   S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, 29, 845-854.
24.   K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, 78, 1950-1958.
25.   A. W. Schuttelkopf and D. M. van Aalten, *Acta Crystallogr D Biol Crystallogr*, 2004, 60, 1355-1363.
26.   P. A. Janne, N. Gray and J. Settleman, *Nat Rev Drug Discov*, 2009, 8, 709-723.
27.   A. Schneider, R. Laage, O. von Ahsen, A. Fischer, M. Rossner, S. Scheek, S. Grunewald, R. Kuner, D. Weber, C. Kruger, B. Klaussner, B. Gotz, H. Hiemisch, D. Newrzella, A. Martin-Villalba, A. Bach and M. Schwaninger, *J Neurochem*, 2004, 88, 1114-1126.
28.   A. P. Kornev, N. M. Haste, S. S. Taylor and L. F. Eyck, *Proc Natl Acad Sci U S A*, 2006, 103, 17783-17788.
29.   J. Lew, *Biochemistry*, 2003, 42, 849-856.
30.   J. M. Murphy, D. M. Korzhnev, D. F. Ceccarelli, D. J. Briant, A. Zarrine-Afsar, F. Sicheri, L. E. Kay and T. Pawson, *Proc Natl Acad Sci U S A*, 2007, 104, 14336-14341.
31.   F. Zuccotto, E. Ardini, E. Casale and M. Angiolini, *J Med Chem*, 2010, 53, 2681-2694.
32.   T. Timm, J. P. von Kries, X. Li, H. Zempel, E. Mandelkow and E. M. Mandelkow, *J Biol Chem*, 2011, 286, 41711-41722.
33.   H. M. Alkahtani, A. Y. Abbas and S. Wang, *Bioorg Med Chem Lett*, 2012, 22, 1317-1321.
34.   Y. Li, J. Geng, Y. Liu, S. Yu and G. Zhao, *ChemMedChem*, 2013, 8, 27-41.

FIGURES AND TABLES

FIGURE LEGEND

Fig. 1: Sequence and structure comparison of human MARK4 kinase domain with MARK 1, 2 and 3 isoforms. (A) Alignment of human MARK1-4 Kinase-UBA sequences. Their functional elements are displayed as colored boxes (Green – P-loop; Black – ATP binding site; Magenta – Hinge region; Blue – Catalytic Loop; Cyan – Activation/T Loop). (B) Predicted structure of human inactive MARK4 Kinase-UBA domain shown in cartoon representation displays the N- [P50-M135] and C- Lobes [E142-E318] of kinase domain in Orange and Wheat color, and UBA [L319-E372] in light green color. Functional elements of kinase domain are specified: P-Loop [K67-A71], Hinge [E136-G141], Catalytic or C-loop [I177-N186] and Activation or T-loop [D199-E225] depicted in same colors as in (A). (C) Superimposition of predicted MARK4-kinase and UBA domains (Deep Purple) with corresponding regions of MARK2 (PDB ID: 1ZMU-A[36] – Light Orange) and MARK3 (PDB ID: 3FE3-A, Blue).

Fig. 2: Molecular dynamics and simulation analysis of MARK4 Kinase-UBA domain. (A) Superimposition of predicted model [magenta] with representative structure of top cluster [green] obtained from the trajectory of 10ns molecular dynamics simulation studies. (B) Backbone RMSD plot depicting the stability of the predicted structure over 10ns timescale. (C) RMSF plot depicting regions of fluctuation across the structure. (D) Principal Component Analysis (PCA) of 10ns trajectory: Projection of first, fifth, tenth and twentieth Eigenvectors obtained from protein coordinate matrix reveals total collective motions of protein and graph denotes the stability of the predicted model over 10ns timescale of MD simulation.

Fig. 3: Structural Comparison of kinase domain in inactive MARK4 and active Aurora Kinase: Superimposition of MARK4 kinase domain – Green, with active Aurora kinase (PDB ID: 1OL5)[37] – Sand color depicts differences in the conformation of αC helix and T loop surrounding the catalytic cleft.

Fig. 4: Unique conformation of DFG motif in MARK kinases. (A) Structural portrait of catalytic cleft observed in inactive Aurora kinase (PDB ID: 1MUO)[38] – Sand color. It shows the presence of DFG-out mode and hydrophobic pocket II (HP-II) in the inactive state. (B) Displays the presence of DFG-in mode and absence of HP-II in active Aurora kinases (PDB ID: 1OL5) – Sand color. The orientation of conserved aspartate (D274) in both modes is highlighted with blue arrowhead. (C) Superimposition of predicted inactive MARK4 DFG motif with that of inactive Aurora kinase highlights the presence of differing orientation of DFG motif of MARK4. (D) Structural comparison of DFG motif in inactive MARK4 and active Aurora kinase confirms that the DFG motif in inactive MARK4 resembles the same orientation (DFG-in) seen in active kinases (Cyan arrowhead depicts the orientation of MARK4 conserved Aspartate, D199.). Figure also depicts the difference in orientation of Glycine residues between inactive MARK4 and active Aurora kinase, suggesting the possibility for occurrence of 'Glycine swing' mechanism resulting in altering the orientation of conserved aspartate. (E) Structural overlay of catalytic cleft in all four isoforms of MARK kinases in inactive state [PDB ID: MARK1: 2HAK[39], MARK2: 2WZJ, MARK3: 3FE3, MARK4: Predicted model] depicts the presence of unusual DFG-in conformation and absence of hydrophobic pocket II adjacent to the catalytic cleft in all inactive MARK kinase domain. (F) Structural

overlay of catalytic cleft between inactive MARK4 model and active MARK2 kinase structure [PDB ID: 3IEC] confirms the presence of intermediate DFG-in conformation in inactive MARK kinase structure.

Fig. 5: Binding analysis of ATP into MARK4 kinase domain. Kinase domain is displayed in cartoon representation with transparent surface. Small black box shows the region of ATP binding surrounded by key structural elements (color coded as per Fig. 1B). Inlay shows interaction between ATP (colored by element) and MARK4 binding residues (H-bonds represented as black dotted lines (Å)).

Fig. 6: Binding mode analysis of known inhibitors with MARK4. (A) Known inhibitors (9-Oxo-derivatives) docked into the ATP-binding pocket of MARK4 kinase domain. The protein is represented as cartoon (N-lobe: yellow; P-loop: red; Hinge: Magenta; C-loop: blue; T-loop: green; C-Lobe: light green; UBA domain: Light Blue) and the binding pocket is shown in surface representation (grey). (B-E) shows interactions of 9-oxo-derivatives- 30199 (sand), 30197 (magenta), 30195 (blue) and 30019 (green) with MARK4 respectively. Ligands and interacting residues are displayed in stick representation colored as per their element. Hydrogen bonds are denoted as black dotted lines; Pi-Pi (Π-Π) interactions in orange, Pi-Sigma (Π-σ) in brown and Pi-Cation (Π-cat+) in green dotted lines. The centroid of aromatic rings in Pi- interactions are displayed as red spheres. Bond distance in Angstrom (Å) units.

Fig. 7: Pharmacophore hypothesis for MARK kinase inhibitors. (A) Two dimensional structure of MARK2 inhibitors (derivatives of 9-oxo-9H-acridin-10-yl: 30019, 30195, 30197, 30199). (B) Pharmacophore mapping from best scored hypothesis (AARRR.9) based on known inhibitors of MARK2. The known inhibitors (30019-green; 30195-blue; 30197-magenta; 30199-sand) are overlaid over the pharmacophore. (C) Pharmacophore features of hypothesis AARRR.9 with a number of sites and their inter-site measurements (Å). Red Balls (A2, A3) represent H-bond acceptors with arrows depicting directional vectors. Rings in Sand color (R7, R8, R9) represent aromatic ring feature in the hypothesis.

Fig. 8: Interactions between MARK4 and top-scoring lead compounds. (A-F) Binding conformation and interaction between residues in ATP binding pocket and compounds 411 (magenta), 214 (green), 950 (blue), 389 (yellow), 781 (forest green) and 529 (orange). Ligands and interacting residues are represented as sticks and are colored as per the elements. (For bond representation and color codes, refer Fig. 5)

Fig. 9: MD simulation analysis of MARK4 complexed with top scoring leads at 10ns timescale. (A) Backbone RMSD plot of MARK4 kinase-UBA domain of ligand complexed structures. (B) RMSD plot of lead compounds in complex with MARK4 depicting changes in ligand binding conformation. (C-H) changes observed in the binding conformation of lead compounds after MD simulation. The native docked conformation (ligand color code: refer Fig. 5) is overlaid onto the conformation obtained post-simulation (ligand color: pink).

Fig. 10: Stability analysis of hydrogen bonds observed between MARK4 and lead compounds in docked and 10ns MD simulated structures. (A-B) depicts hydrogen bond distance profile of MARK4-411; (C) MARK4-214; (D-E) MARK4-950; (F) MARK4-389; (G) MARK4-781; (H) MARK4-529 complex.

Fig. 11: Analysis of the changes occurred in ATP-Binding site of MARK4: (A) details the changes observed in the binding site volume of MARK4 inhibitors. The binding site volume is calculated for native-unbound form and for both docked and simulated complexes; values are compared (brown-docked; sand – simulated complex) and displayed. (B-G) displays the differences observed in the volume and area of MARK4 inhibitors binding site. In each segment, the left side displays binding site of docked complex whereas the right side displays that of simulated complex. Based on the results and structural representation, we can observe that for all complexes, except Lig-950, the conformation of binding pocket is altered and has closed in after ligand binding (For color codes of ligands refer Fig. 6).

Fig. 12: Workflow: Towards identification of novel MARK4 inhibitors.

TABLES

Table 1: Docking Score and interaction analysis for known inhibitors with MARK2 and MARK4

Title	XP Glide Score Kcal/mol	Glide Energy Kcal/mol	H-Bond Interactions		Pi Interactions			Charge Interaction	Van der Waals' Interacting Residues
			Donor-Acceptor	Dist. (Å)	Residues	Type	Dist. (Å)		
MARK2									
9-oxo-30195	-6.685	-44.588	A132:H - UNK:O1	1.93	Y131 Y131	Π-Π Π-Π	5.78 5.99	I59,K61, N198, E199	G60,V67,A80,M129, E130,G135,E136,E179, N180,L182
9-oxo-30197	-6.656	-46.281	A132:H -UNK:O1	1.92	Y131 Y131	Π-Π Π-Π	5.78 6.00	I59,K61, E199	G60,V67,A80,M129, E130,G135,E136,K177, E179,N198
9-oxo-30199	-6.350	-47.743	A132:H - UNK:O1	1.95	K61:NZ Y131 Y131	Π-cat+ Π-Π Π-Π	3.92 5.61 5.72	E136,E199	I59,G60,V67,A80,M129,E1 30,S133,G135,K177, E179,N198,T208,F209, C210
9-oxo-30019	-4.990	-44.896	A132:H - UNK:O1	2.06	E199:HB3	Π-σ	2.80	K61,N198	I59,V67,A80,V113, M129,E130,S133,G135, E136,K177,E179,N180, L182,A192,D193,C210
MARK4									
9-oxo-30199	-6.842	-48.111	K88:HZ1 - UNK:O2 K88:HZ1 - UNK:N3 A138:H - UNK:O1	2.44 2.28 1.98	K88:NZ Y137 D199:HB3	Π-cat+ Π-Π Π-σ	3.91 5.86 2.44	M110,L208 ,T207	I65,G66,V73,A86,L103, E106,V119,L133,M135, E136,Y137,S139,L188, A198,F206
9-oxo-30197	-5.834	-42.761	A138:H - UNK:O1	1.91	Y137 Y137	Π-Π Π-Π	5.61 6.21	E136,E185, D199	I65,V73,A86,K88,V119,M 135,Y137,S139,G141,E142 ,L188,A198,L208
9-oxo-	-5.044	-42.599	K88:HZ1 - UNK:O2	1.75	K88:NZ	Π-cat+	4.18	D199,T207	I65,V73,A86,E106,M110,V

| 30195 | | | | | K88:NZ
K88:NZ
K88:HE3
F202 | Π-cat+
Π-cat+
Π-σ
Π-Π | 5.91
6.01
5.40
2.43 | ,L208 | 119,M135,E136,Y137,E14
2,E185,L188,A198,F202,S
203,F206 |
| 9-oxo-
30019 | -4.937 | -39.399 | K88:HZ1 - UNK:O2 | 1.79 | F202 | Π-Π | 5.35 | E106,M110
,A138,D19
9,T207,L20
8 | I65,V73,A86,V119,M135,E
136,Y137,E142,E185,L188
,A198,F202,F206 |

Table 2: Statistical results of PHASE generated Pharmacophore hypothesis for MARK4 Kinase inhibitors

ID	Survival	Site	Vector	Volume	Selectivity	No. of Matches	Activity (-Plog)	Cluster Number	Cluster Size	Average Similarity
AARRR.9	3.357	0.89	0.991	0.481	1.499	3	-0.146	1	2	1.000
AARRR.18	3.357	0.89	0.991	0.481	1.499	3	-0.146	1	2	1.000
AARRR.4	3.253	0.77	0.977	0.503	1.505	3	-0.114	2	4	1.000
AARRR.7	3.253	0.77	0.977	0.503	1.505	3	-0.114	2	4	1.000
AARRR.13	3.253	0.77	0.977	0.503	1.505	3	-0.114	2	4	1.000
AARRR.16	3.253	0.77	0.977	0.503	1.505	3	-0.114	2	4	1.000
ARRRR.21	2.776	0.54	0.711	0.529	1.684	3	-0.204	3	1	0.000

A- H-bond Acceptor; R-Aromatic Ring
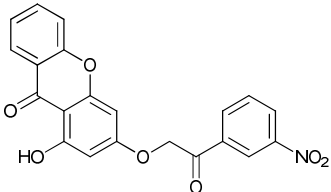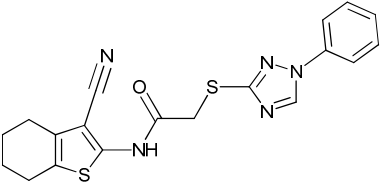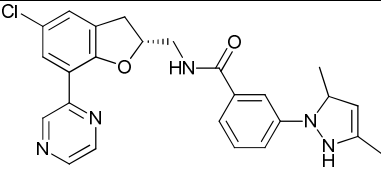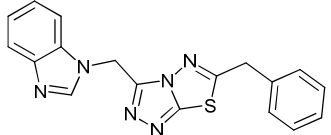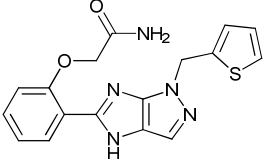
Table 3: Alignment of known inhibitors for best scored hypothesis AARRR.9

Ligand Name	Activity (-Plog)	Pharm Set	Fitness	# Sites Matched	Relative Energy
9-oxo-30019	-0.146	active	3.00	5	0.000
9-oxo-30195	-0.114	active	2.27	5	0.000
9-oxo-30199	-0.204	active	2.45	5	0.000
9-oxo-30197	-0.544	inactive	-	-	-

Table 4: Inter-site distances between Pharmacophore features of hypothesis AARRR.9

Entry	Site1	Site2	Distance (Å)
AARRR.9	R7	R8	4.943
AARRR.9	A2	A3	4.738
AARRR.9	A3	R8	4.145
AARRR.9	A3	R7	4.142
AARRR.9	A2	R7	3.701
AARRR.9	A2	R8	3.701
AARRR.9	A3	R9	3.173
AARRR.9	A2	R9	2.708
AARRR.9	R7	R9	2.479
AARRR.9	R8	R9	2.476

6

Table 5: Top scoring hits matched according to best scoring hypothesis AARRR.9

S. No.	2D Structure	Entry ID	Chembridge ID	Mol. Wt. (kD)	H-Acc.	H-Don	Align Score	Fitness	Vector Score	Volume Score	Matched Ligand Sites
1.		411	7958232	391.336	7	1	1.0009	1.2709	0.8083	0.2967	A(2) A(4) R(8) R(10) R(7)
2.		214	7749255	395.496	5	1	1.3958	0.6461	0.4782	0.3310	A(2) A(1) R(8) R(10) R(9)
3.		950	88650970	459.934	6	1	1.5631	0.4818	0.4408	0.3435	A(4) A(1) R(12) R(14) R(13)
4.		389	7955248	346.409	4	0	1.1924	0.9638	0.5479	0.4095	A(4) A(1) R(6) R(8) R(7)
5.		781	59093313	353.398	4	2	1.2388	1.0277	0.6646	0.3955	A(1) A(2) R(9) R(11) R(10)

| 6. | | 529 | 19240599 | 348.363 | 5 | 2 | 1.0953 | 0.8987 | 0.5247 | 0.2868 | A(1) A(4) R(11) R(10) R(9) |

H-Acc. – Hydrogen bond acceptor; H-Don. – Hydrogen bond donor.

Table 6: Docking Score and interaction analysis for top scoring hits with MARK4

Ligand	XP Glide Score Kcal/mol	Glide Energy Kcal/mol	H-Bond Interactions		Pi Interactions			Charge Interaction	Van der Waals' Interacting Residues
			Donor-Acceptor	Dist. (Å)	Residues	Type	Dist. (Å)		
411	-8.453	-48.247	UNK:H6 - E136:O A138:H - UNK:O7	2.08 2.03	-	-	-	K88,M135,Y137,D199,T207,F206,E106	G66,V73,A86,M110,V119,G141,A198,F202,L208
214	-8.067	-49.394	D199:H - UNK:N1	2.23	K88:NZ M135:HE2 Y137 F206	Π-cat+ Π-σ Π-Π Π-Π	5.30 2.87 5.74 6.00	M110, E136,A138	I65,V73, A86, I90, L103, E106, V119,L133, S139,G141,L188,A198, F202,L208
950	-7.386	-34.997	K88:HZ1 - UNK:O2 A138:H - UNK:N4 UNK:H3 - D199:OD1	2.19 2.25 2.40	K88:NZ F202	Π-cat+ Π-Π	4.95 6.04	M110,E136,T207,L208	I65,V73,A86,I90,L103,V107,E106,V119,L133,M135,Y137,E142,E185,L188,A198,S203,F206
389	-7.290	-41.493	K88:HZ1 - UNK:S1 K88:HZ1 - UNK:N2 A138:H - UNK:N5	2.42 2.48 2.26	K88:NZ K88:NZ K88:NZ	Π-cat+ Π-cat+ Π-cat+	3.57 3.41 4.80	E136,T207,L208	I65,V73,A86,I90,L103,E106,M110,V119,L133,M135,Y137,L188,D199, F202,S203,F206
781	-7.137	-47.241	UNK:H34 - E185:O UNK:H35 - D199:OD1	1.77 1.83	-	-	-	K88,E106,M110,N186,T207	I65,V73,A86,I90,L103,V107,V119,M135,E136,Y137,A138,L188,F202,F206,L208
529	-7.126	-42.473	A138:H - UNK:N42 UNK:H26 - D199:OD1	2.14 1.88	K88:NZ F202	Π-cat+ Π-Π	3.89 5.71	K88,E136,T207	I65,V73,A86,M110,V119,M135,Y137,E185,L188,A198,F202,S203,F206,L208

Fig. 1: Sequence and structure comparison of human MARK4 kinase domain with MARK 1, 2 and 3 isoforms. (A) Alignment of human MARK1-4 Kinase-UBA sequences. Their functional elements are displayed as colored boxes (Green – P-loop; Black – ATP binding site; Magenta – Hinge region; Blue – Catalytic Loop; Cyan – Activation/T Loop). (B) Predicted structure of human inactive MARK4 Kinase-UBA domain shown in cartoon representation displays the N- [P50-M135] and C- Lobes [E142-E318] of kinase domain in Orange and Wheat color, and UBA [L319-E372] in light green color. Functional elements of kinase domain are specified: P-Loop [K67-A71], Hinge [E136-G141], Catalytic or C-loop [I177-N186] and Activation or T-loop [D199-E225] depicted in same colors as in (A). (C) Superimposition of predicted MARK4-kinase and UBA domains (Deep Purple) with corresponding regions of MARK2 (PDB ID: 1ZMU-A36 – Light Orange) and MARK3 (PDB ID: 3FE3-A, Blue).
173x193mm (300 x 300 DPI)

Fig. 2: Molecular dynamics and simulation analysis of MARK4 Kinase-UBA domain. (A) Superimposition of predicted model [magenta] with representative structure of top cluster [green] obtained from the trajectory of 10ns molecular dynamics simulation studies. (B) Backbone RMSD plot depicting the stability of the predicted structure over 10ns timescale. (C) RMSF plot depicting regions of fluctuation across the structure. (D) Principal Component Analysis (PCA) of 10ns trajectory: Projection of first, fifth, tenth and twentieth Eigen vectors obtained from protein coordinate matrix reveals total collective motions of protein and graph denotes the stability of the predicted model over 10ns timescale of MD simulation.
173x230mm (300 x 300 DPI)
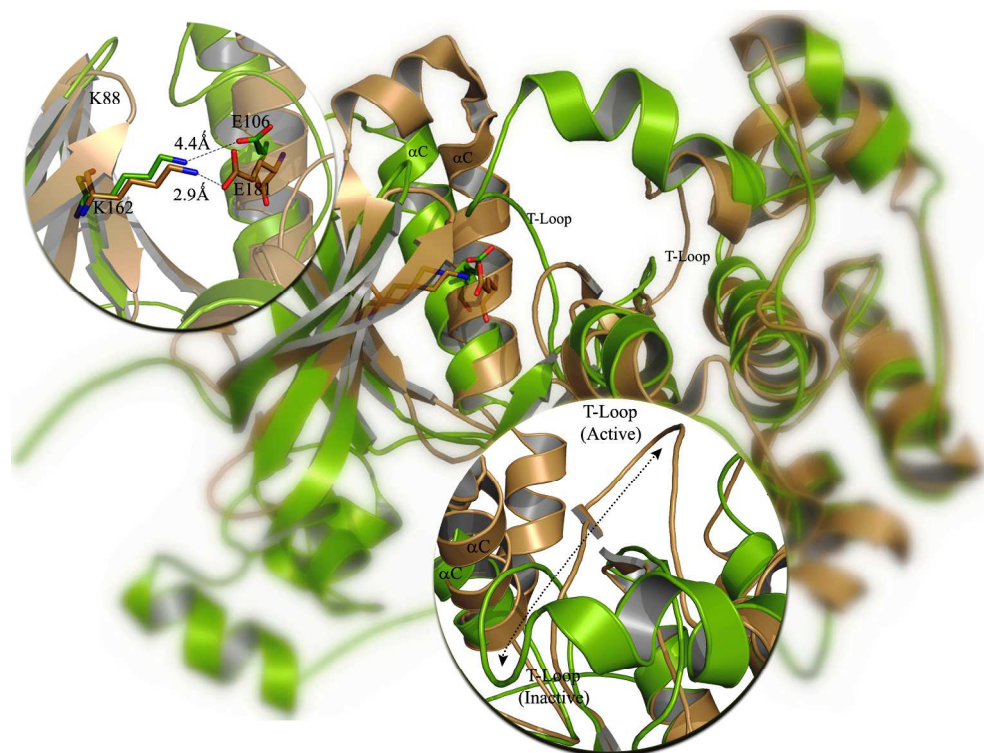
Fig. 3: Structural comparison of kinase domain in inactive MARK4 and active Aurora Kinase: Superimposition of MARK4 kinase domain – Green, with active Aurora kinase (PDB ID: 1OL5)37 – Sand color depicts differences in the conformation of αC helix and T loop surrounding the catalytic cleft.
230x173mm (300 x 300 DPI)

Fig. 4: Unique conformation of DFG motif in MARK kinases. (A) Structural portrait of catalytic cleft observed in inactive Aurora kinase (PDB ID: 1MUO)38 – Sand color. It shows the presence of DFG-out mode and hydrophobic pocket II (HP-II) in the inactive state. (B) Displays the presence of DFG-in mode and absence of HP-II in active Aurora kinases (PDB ID: 1OL5) – Sand color. The orientation of conserved aspartate (D274) in both modes is highlighted with blue arrowhead. (C) Superimposition of predicted inactive MARK4 DFG motif with that of inactive Aurora kinase highlights the presence of differing orientation of DFG motif of MARK4. (D) Structural comparison of DFG motif in inactive MARK4 and active Aurora kinase confirms that the DFG motif in inactive MARK4 resembles the same orientation (DFG-in) seen in active kinases (Cyan arrowhead depicts the orientation of MARK4 conserved aspartate, D199.). Figure also depicts the difference in orientation of glycine residues between inactive MARK4 and active Aurora kinase, suggesting the possibility for occurrence of 'Glycine swing' mechanism resulting in altering the orientation of conserved aspartate. (E) Structural overlay of catalytic cleft in all four isoforms of MARK kinases in inactive state [PDB ID: MARK1: 2HAK39, MARK2: 2WZJ, MARK3: 3FE3, MARK4: Predicted model] depicts the presence of unusual DFG-in conformation and absence of hydrophobic pocket II adjacent to the catalytic cleft in all inactive MARK kinase domain. (F) Structural overlay of catalytic cleft between inactive MARK4 model and active MARK2 kinase structure [PDB ID: 3IEC] confirms the presence of intermediate DFG-in conformation in inactive MARK kinase structure.
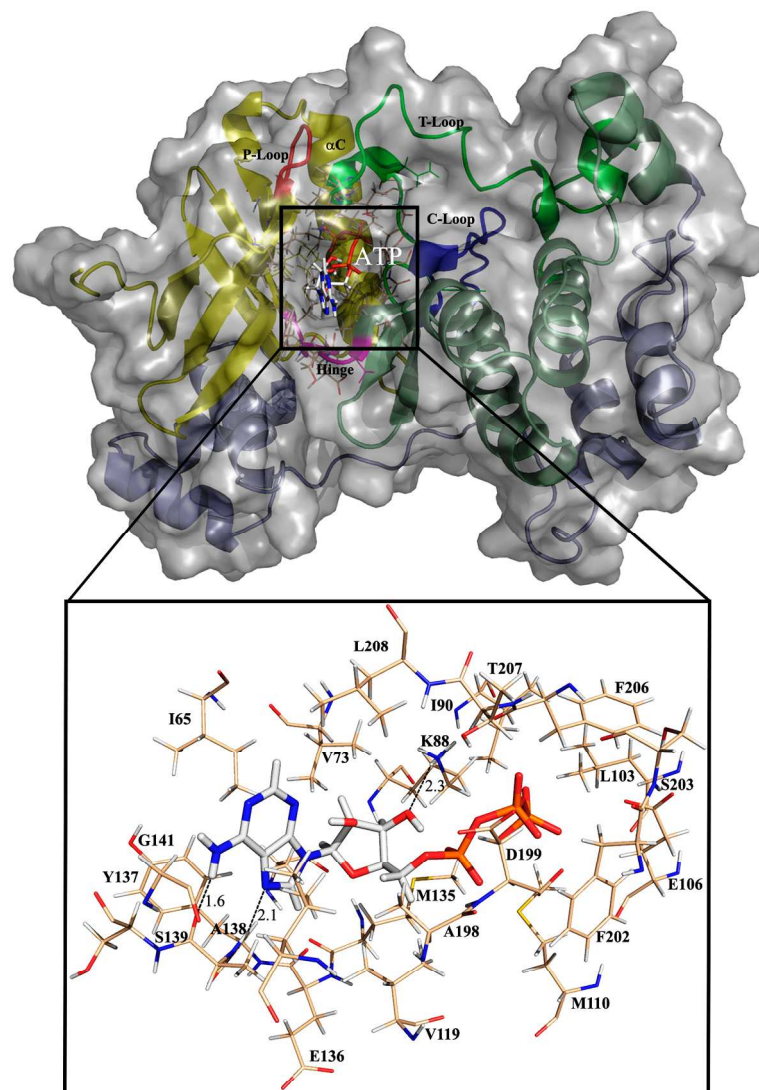165x171mm (300 x 300 DPI)

Fig. 5: Binding analysis of ATP into MARK4 kinase domain. Kinase domain is displayed in cartoon representation with transparent surface. Small black box shows the region of ATP binding surrounded by key structural elements (color coded as per Fig. 1B). Inlay shows interaction between ATP (colored by element) and MARK4 binding residues (H-bonds represented as black dotted lines (Å)).
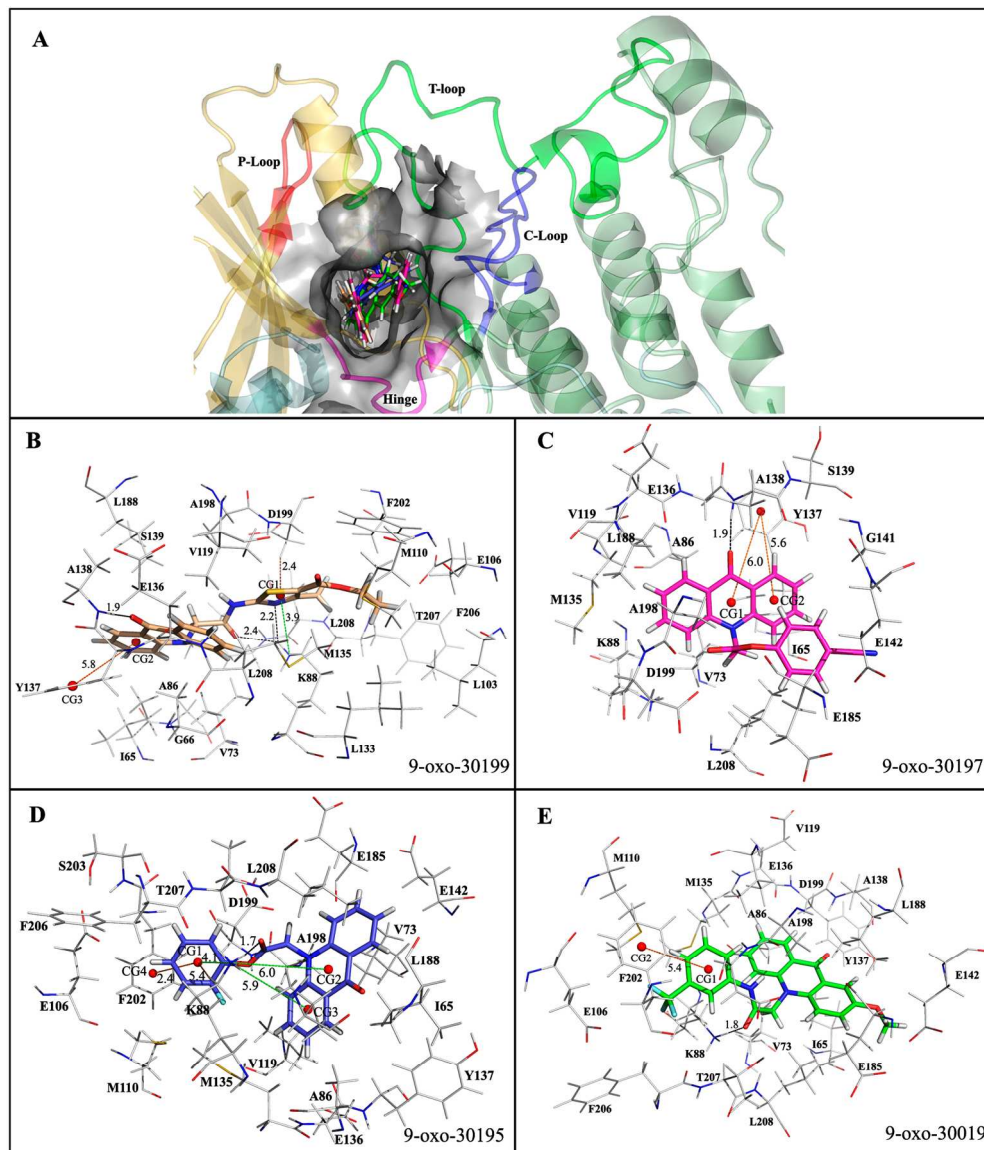173x230mm (300 x 300 DPI)

Fig. 6: Binding mode analysis of known inhibitors with MARK4. (A) Known inhibitors (9-Oxo-derivatives) docked into the ATP-binding pocket of MARK4 kinase domain. The protein is represented as cartoon (N-lobe: yellow; P-loop: red; Hinge: Magenta; C-loop: blue; T-loop: green; C-Lobe: light green; UBA domain: Light Blue) and the binding pocket is shown in surface representation (grey). (B-E) shows interactions of 9-oxo-derivatives- 30199 (sand), 30197 (magenta), 30195 (blue) and 30019 (green) with MARK4 respectively. Ligands and interacting residues are displayed in stick representation colored as per their element. Hydrogen bonds are denoted as black dotted lines; Pi-Pi (Π-Π) interactions in orange, Pi-Sigma (Π-σ) in brown and Pi-Cation (Π-cat+) in green dotted lines. The centroid of aromatic rings in Pi- interactions are displayed as red spheres. Bond distance in Angstrom (Å) units.
162x189mm (300 x 300 DPI)
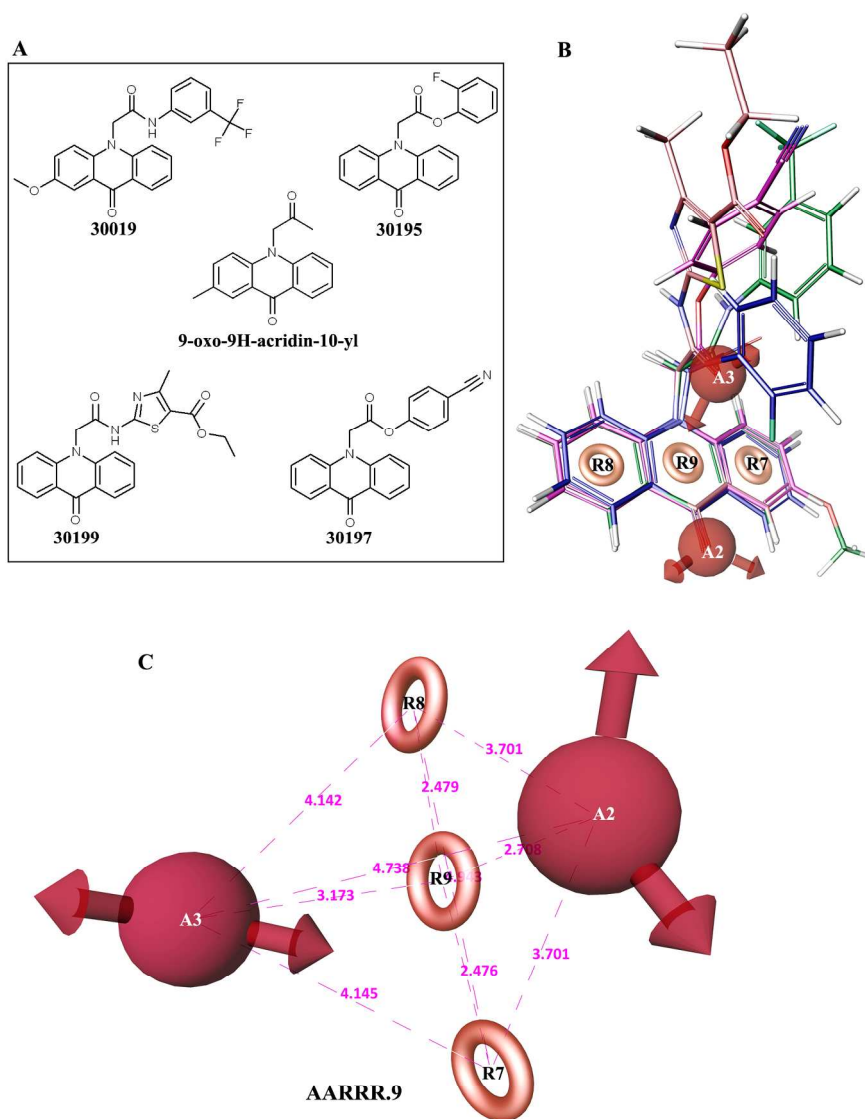
Fig. 7: Pharmacophore hypothesis for MARK kinase inhibitors. (A) Two dimensional structure of MARK2 inhibitors (derivatives of 9-Oxo-9H-acridin-10-yl: 30019, 30195, 30197, 30199). (B) Pharmacophore mapping from best scored hypothesis (AARRR.9) based on known inhibitors of MARK2. The known inhibitors (30019-green; 30195-blue; 30197-magenta; 30199-sand) are overlaid over the pharmacophore. (C) Pharmacophore features of hypothesis AARRR.9 with a number of sites and their inter-site measurements (Å). Red Balls (A2, A3) represent H-bond acceptors with arrows depicting directional vectors. Rings in Sand color (R7, R8, R9) represent aromatic ring feature in the hypothesis.
173x230mm (300 x 300 DPI)
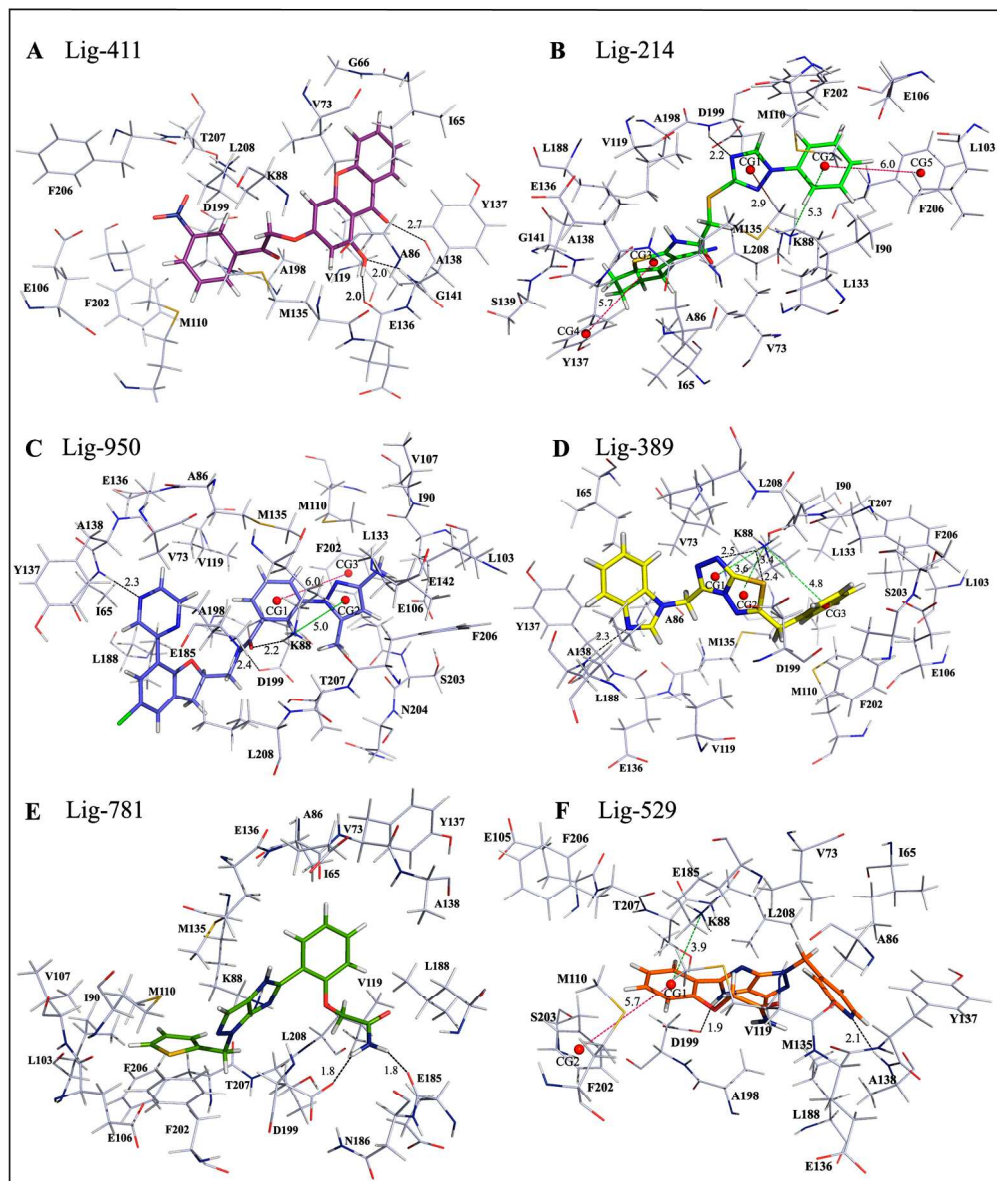
Fig. 8: Interactions between MARK4 and top-scoring lead compounds. (A-F) Binding conformation and interaction between residues in ATP binding pocket and compounds 411 (magenta), 214 (green), 950 (blue), 389 (yellow), 781 (forest green) and 529 (orange). Ligands and interacting residues are represented as sticks and are colored as per the elements. (For bond representation and color codes, refer Fig. 5)
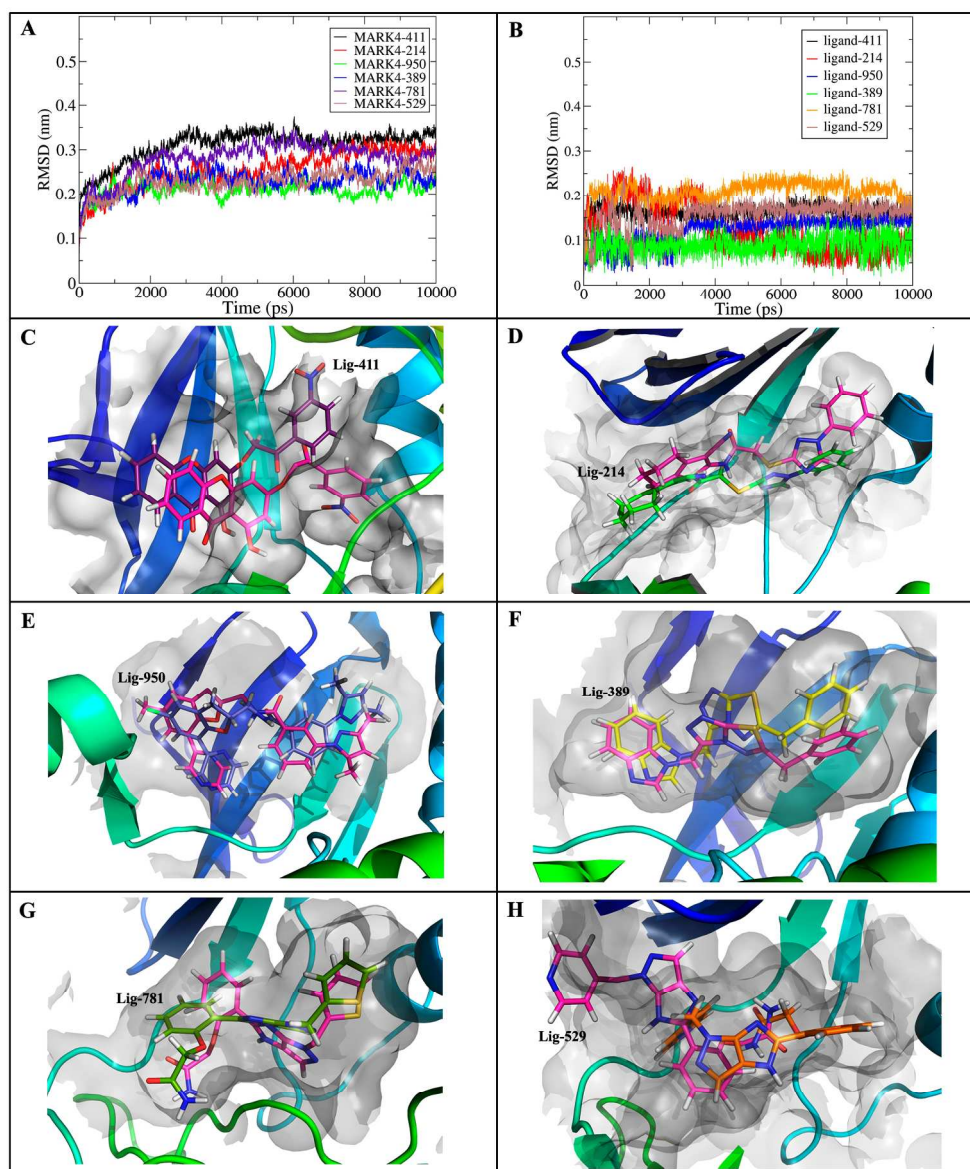173x205mm (300 x 300 DPI)

Fig. 9: MD simulation analysis of MARK4 complexed with top scoring leads at 10ns timescale. (A) Backbone RMSD plot of MARK4 kinase-UBA domain of ligand complexed structures. (B) RMSD plot of lead compounds in complex with MARK4 depicting changes in ligand binding conformation. (C-H) changes observed in the binding conformation of lead compounds after MD simulation. The native docked conformation (ligand color code: refer Fig. 5) is overlaid onto the conformation obtained post-simulation (ligand color: pink). 193x230mm (300 x 300 DPI)
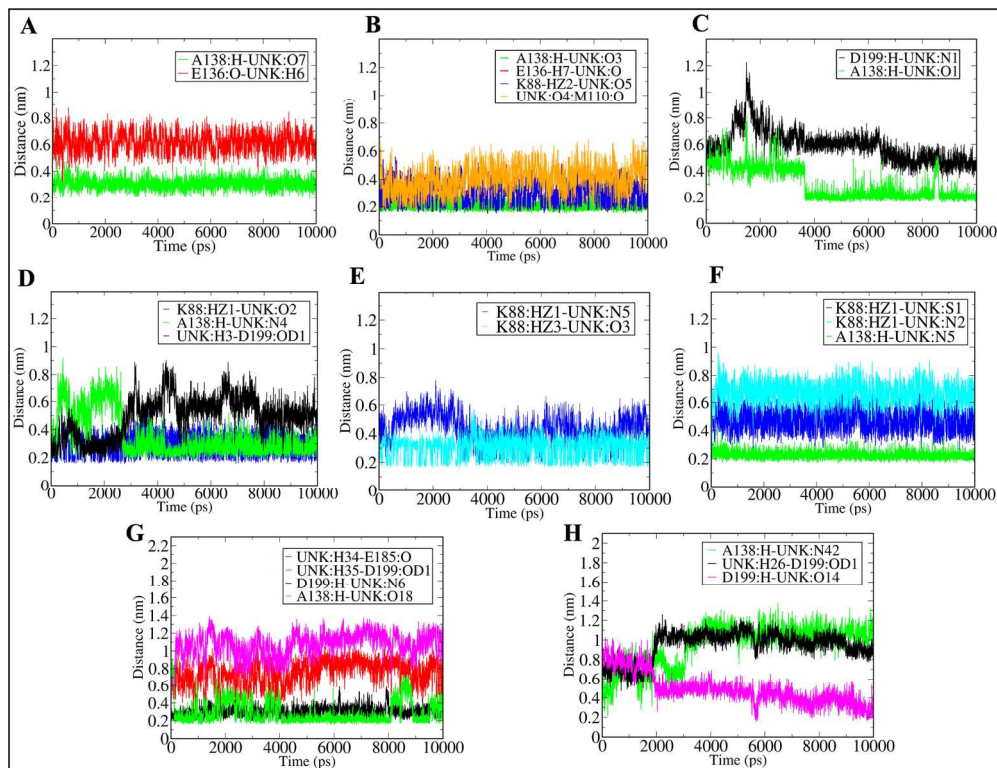
Fig. 10: Stability analysis of hydrogen bonds observed between MARK4 and lead compounds in docked and 10ns MD simulated structures. (A-B) depicts hydrogen bond distance profile of MARK4-411; (C) MARK4-214; (D-E) MARK4-950; (F) MARK4-389; (G) MARK4-781; (H) MARK4-529 complex.
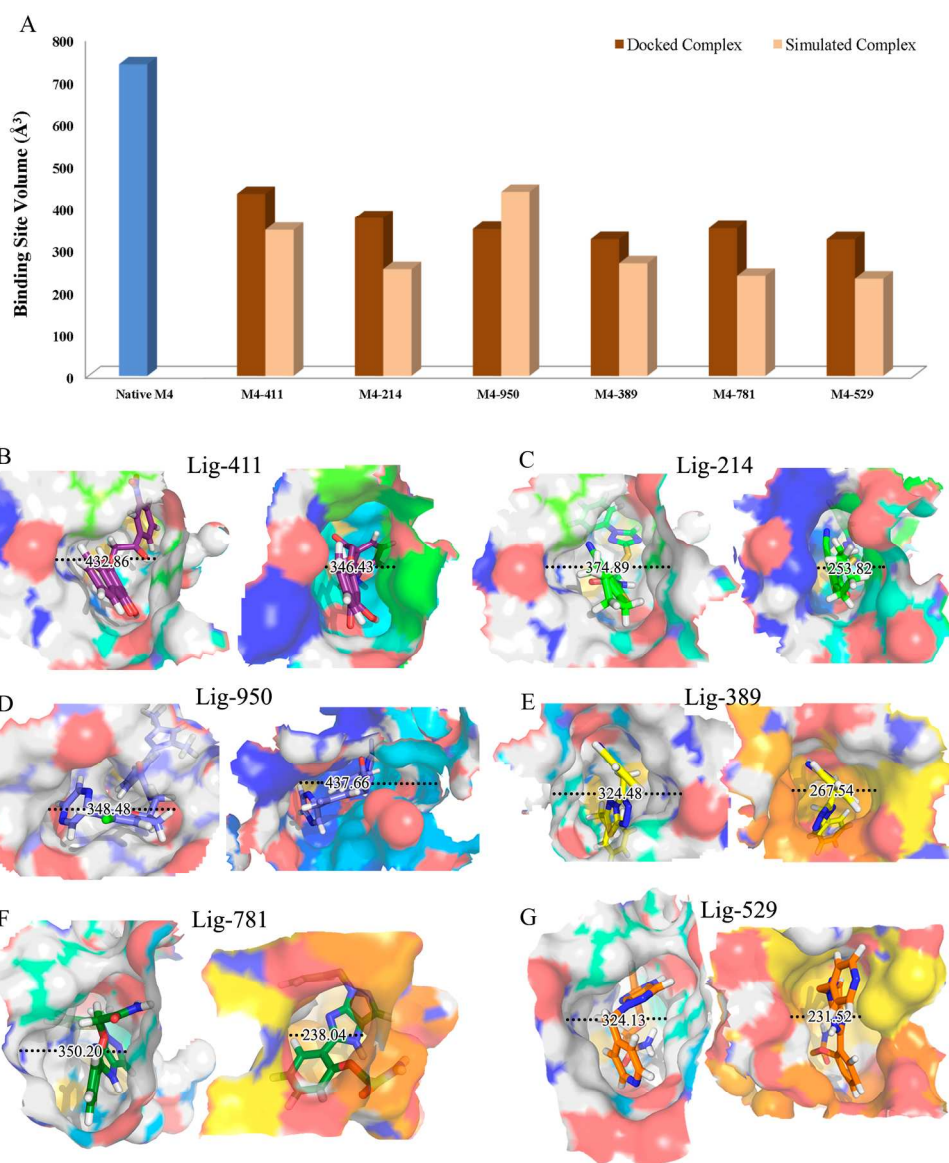173x134mm (300 x 300 DPI)

Fig. 11: Analysis of the changes occurred in ATP-Binding site of MARK4: (A) details the changes observed in the binding site volume of MARK4 inhibitors. The binding site volume is calculated for native-unbound form and for both docked and simulated complexes; values are compared (brown-docked; sand –simulated complex) and displayed. (B-G) displays the differences observed in the volume and area of MARK4 inhibitors binding site. In each segment, the left side displays binding site of docked complex whereas the right side displays that of simulated complex. Based on the results and structural representation, we can observe that for all complexes, except Lig-950, the conformation of binding pocket is altered and has closed in after ligand binding (For color codes of ligands refer Fig. 6).
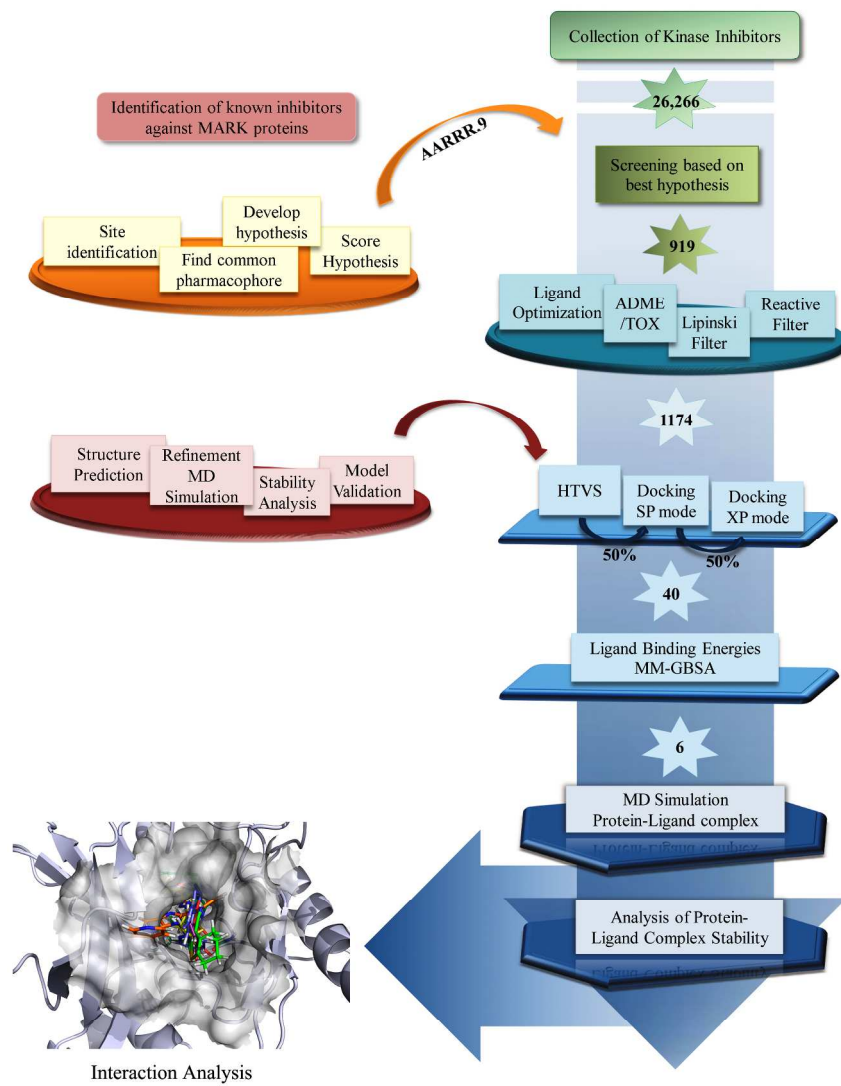173x212mm (300 x 300 DPI)

Fig. 12: Workflow: Towards identification of novel MARK4 inhibitors.
209x297mm (300 x 300 DPI)