

Targeting Plague Virulence Factors: A Combined Machine Learning Method and Multiple Conformational Virtual Screening for the Discovery of *Yersinia* Protein Kinase A Inhibitors

Xin Hu,[†] Gerd Prehna,[†] and C. Erec Stebbins*

Laboratory of Structural Microbiology, The Rockefeller University,
New York, New York 10021

Received June 5, 2007

Abstract: *Yersinia* spp. is currently an antibiotic resistance concern and a re-emerging disease. The essential virulence factor *Yersinia* protein kinase A (YpkA) contains a Ser/Thr kinase domain whose activity modulates pathogenicity. Here, we present an approach integrating a machine learning method, homology modeling, and multiple conformational high-throughput docking for the discovery of YpkA inhibitors. These first reported inhibitors of YpkA may facilitate studies of the pathogenic mechanism of YpkA and serve as a starting point for development of anti-plague drugs.

Many Gram-negative bacterial pathogens utilize a type III secretion system (TTSS) to inject effector proteins into the cytosol of host cells.¹ These virulence factors play an important role in bacterial pathogenesis by modulating the host processes that regulate actin cytoskeletal assembly.² With the emergence of antibiotic resistance and the threat of such bacteria being used as biological weapons, targeting virulence proteins for antibiotic design is attractive, as such compounds are unlikely to be cross-resistant or to induce resistance.^{3,4}

Herein, we report our efforts on the discovery of inhibitors for the *Yersinia* protein kinase A (YpkA^a). YpkA is an essential virulence determinant in *Yersinia* spp., which includes the causative agent of plague.⁵ The protein contains a chaperone binding/membrane localization domain,⁶ a Ser/Thr kinase domain, a GDI-like domain that interacts with the Rho family of small GTPases, and a C-terminal subdomain responsible in part for actin binding and kinase activation.^{7,8} Because the kinase activity of YpkA has been shown to directly correlate to virulence by phosphorylating the small G protein Gαq, inhibition of YpkA could yield new antiplague therapeutics.^{9,10}

Protein kinase inhibitor design is a challenging problem because of the high similarity and plasticity of the catalytic site.^{11–14} In this study, we applied an approach combining a machine learning method and multiple conformational high-throughput docking for the discovery of YpkA inhibitors. The screening strategy employed is illustrated in Figure 1. First, we developed a machine learning support vector machine (SVM) model using a data set of known kinase inhibitors from a diverse kinase collection. The ligand-based SVM model was used as a kinase filter to prioritize the large size of chemical databases, and a target-focused library was obtained. Second, we constructed homology models of YpkA based on the MAPK templates and further performed MD simulations to sample

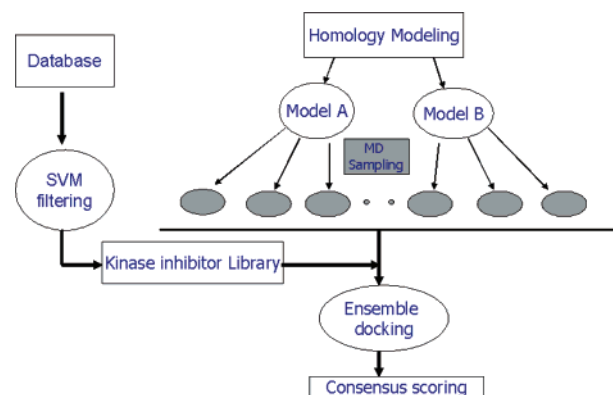


Figure 1. Virtual screening strategy combined machine learning method, homology modeling, and multiple conformational high-throughput docking for the discovery of YpkA inhibitors.

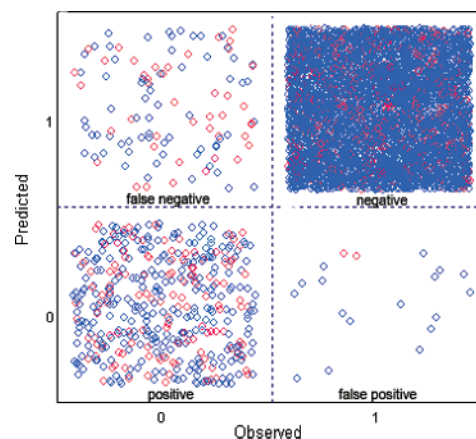


Figure 2. Machine learning SVM model derived from kinase inhibitors. The “positives” are known kinase inhibitors and the “negatives” are randomly selected, inactive compounds. The “false negatives” and “false positives” are those that were predicted incorrectly by the SVM model.

different protein conformations characterized in the catalytic site to account for protein flexibility. Finally, with an ensemble of protein structures and the kinase inhibitor-enriched library, multiple conformational high-throughput docking was performed and a number of potent and selective inhibitors of YpkA have been successfully identified.

To develop a general kinase model for database filtering, a data set of known kinase inhibitors was seeded into a randomly selected chemical library serving as the training set. These kinase inhibitors were assembled from the literature and those in complex with a protein target deposited in the Protein Data Bank (PDB). In total, 364 kinase inhibitors were selected covering a diversity of known chemical scaffolds for both Ser/Thr kinases and Tyr kinases. The inactive data set comprising 4220 compounds was randomly selected from the MDDR database (Elsevier MDL, San Leandro, CA). Molecular descriptors were calculated with ADMET/Predictor consisting of 276 descriptors from the 3D structure (SimulationsPlus, Lancaster, CA). The use of ADMET molecular descriptors was anticipated to improve the druglikeness property of identified compounds, which is a crucial aspect in the late stage of drug development. The SVM model was derived from the molecular descriptors of the training set in distinguishing the active and inactive compounds. As shown in Figure 2, 319 out of 364 inhibitors were classified in the “positive” region, while only 15% of the

* To whom correspondence should be addressed. Phone: 212-327-7190. Fax: 212-327-9171. E-mail: stebbins@rockefeller.edu.

[†] These authors contributed equally to this work.

^a Abbreviations: G-loop, glycine loop; GDI, guanine nucleotide dissociation inhibitor; MD, molecular dynamics; PKC, protein kinase C; PDB, Protein Data Bank; SVM, support vector machine; YpkA, *Yersinia* protein kinase A.

active compounds were misclassified as false negatives. To validate the model, a testing data set comprising 175 known kinase inhibitors and 669 inactive compounds was applied using the SVM model. 127 out of 175 active compounds were predicted correctly, yielding an enrichment of 70%. This result is promising and comparable to many other machine learning models published recently.^{15,16} Given the high efficiency of the SVM model, we then screened our in-house database collections consisting of more than 2 million compounds, and a kinase-focused library of ~200 000 compounds was obtained.

Because the structure of YpkA is unavailable, we constructed 3D models based on MAPK. YpkA shares about 20% homology to mammalian Ser/Thr kinases (Figure 3), but considering only the residues near the ATP binding site, the sequence identity to MAPK is 60%. Therefore, there is enough similarity to build a reliable model focused on the catalytic site. Two structural models were constructed on the basis of different templates of MAPK. Model A used the apo structures of p38 (PDB codes 1p38 and 1erk), while model B adopted ligand-bound complexes with induced fit at the ATP binding site (PDB codes 1a9u and 3erk). As shown in Figure 3, structural differences can be seen within these two models. Model A possesses a more open ATP binding pocket at the Glycine loop (G-loop), while the catalytic site in model B is closed with the G-loop flipping down. Because the conformational changes of the G-loop are sensitive to ligand perturbation, both are valid conformations for inhibitor design. We also examined the DFG motif, which is a key element in kinase inhibitor design.¹⁷ Although in YpkA the corresponding motif is DLG, His293 following the DLGL motif could potentially act in a similar manner to the Phe in mammalian kinases, namely, “His-in” and “His-out” as modeled in A and B.

To further examine the structural features of YpkA, we performed molecular dynamic simulations using both the apo and ATP bound models. The simulations were carried out in vacuo to permit more extended conformational changes. Analysis of the dynamics of the protein at different states revealed a number of active site residues that exhibited high flexibility (Figure 3). To sample a good representation of protein conformations for the subsequent ensemble docking, 500 conformers were extracted from 2.0 ns MD simulations and clustered according to a defined residue center at the active site. Five major clusters were obtained with model A and three clusters with model B. From the MD simulations and the docking studies we believe that the conformational changes of the active site residues represent to some extent the plasticity of the ATP binding site upon ligand binding.

With the model of YpkA and the SVM-enriched kinase inhibitor library, we then performed a multiple conformational high-throughput docking to search for YpkA inhibitors. The program FlexE was used, which accommodates multiple protein conformations in docking by forming new structural representatives.¹⁸ A total of eight conformers of YpkA sampled from MD simulations were used in FlexE docking. A kinase focused library consisting of ~200 000 compounds was subsequently docked into the ensemble of protein structures and ranked according to the FlexX score. To improve the hit selection, we also applied consensus scoring for the FlexX-docked complexes. The top 5% compounds were extracted and reranked using X-Score.¹⁹ The top 1000 compounds from both the original FlexX score and the X-Score were visually inspected in terms of overall fit, interactions in the binding site, as well as structural complexity and diversity. A total of 45 compounds were finally selected to experimentally test against YpkA. Seven of the 45

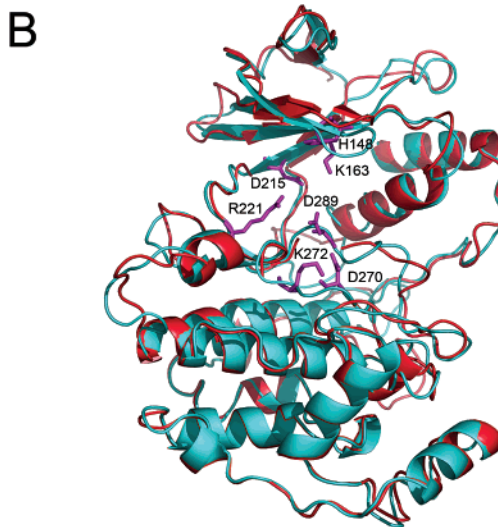
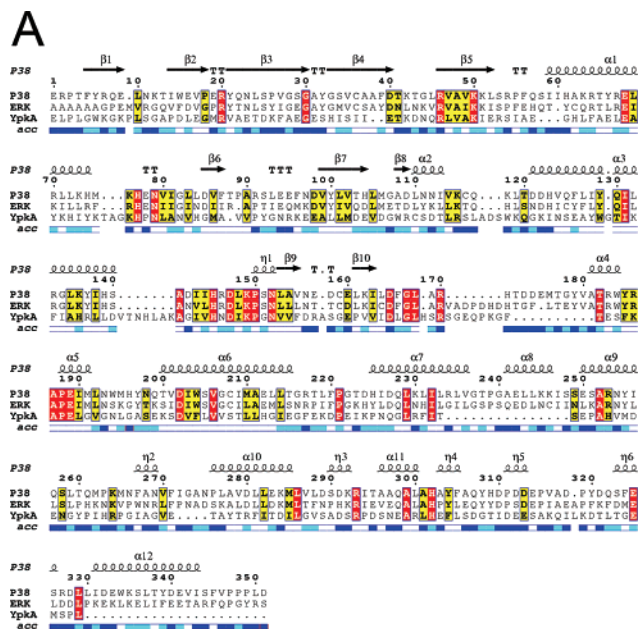


Figure 3. (A) Sequence alignment of YpkA (115–431) with protein kinases P38 and ERK. Strict sequence conservation is shown in red, and strong sequence conservation is shown in yellow. The solvent accessibility of each residue in the P38 structure is indicated in the bar at the base of the sequences, with white representing buried residues, dark-blue representing solvent-accessible residues, and light-blue representing an intermediate value. The secondary structural elements are also indicated according to the structure of P38. (B) Structural alignment of the two homology models of YpkA kinase domain. Model A (red) represents a conformation of YpkA with an open ATP-binding pocket, while model B (cyan) has a closed ATP-binding pocket. The key residues to the ligand binding are shown in magenta.

compounds showed complete inhibition at 225–450 μM , yielding a hit rate of 15%. The IC_{50} values of these compounds were determined by radiological assay with three compounds exhibiting inhibitory activities at 1.81, 5.87, and 9.72 μM and the remaining four having IC_{50} values below 50 μM . Examination of these active compounds revealed a diversity of chemical structure, as represented in Figure 4. Compound **1** possesses a scaffold of indolin-2, which is found in the derivatives of CDK2 inhibitors.²⁰ Compound **2** belongs to the anthraquinone family, potent inhibitors of casein kinase-2.²¹ The structures of **3** and **4** are quite interesting, as they possess the functional group pyrimidine-2,4,6-trione. As can be seen in Figure 5, the binding

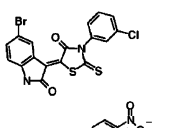
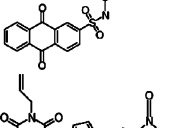
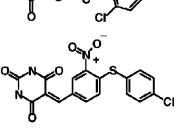
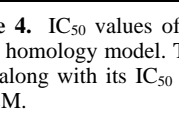
Compound	IC ₅₀ (μM)		
	YpkA	MAPK	PKC
 1	1.81 ± 0.33	2.45 ± 0.53	> 200
 2	19.84 ± 4.32	> 200	165 ± 61
 3	9.72 ± 1.66	54.46 ± 19.4	205.16 ± 47.65
 4	5.87 ± 0.81	63 ± 21	> 200

Figure 4. IC₅₀ values of inhibitors from virtual screening with the kinase homology model. The chemical structure of each compound is listed along with its IC₅₀ value measured at YpkA concentrations of 0.15 μM.

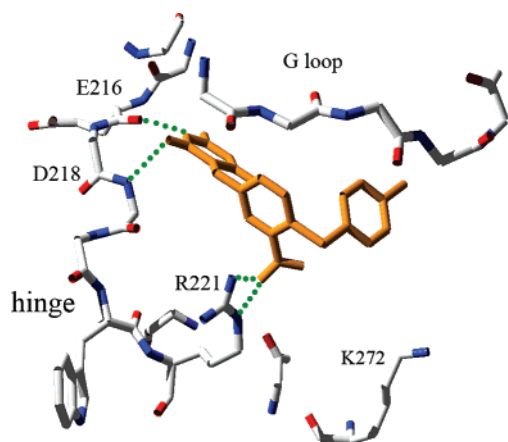


Figure 5. Predicted intermolecular interactions of **4** in the YpkA active site. The YpkA inhibitor is shown in orange, hydrogen-bonding interactions are shown in green, and the relevant residues of YpkA are colored by atom type and labeled.

mode of pyrimidine derivative in YpkA resembles the adenosine moiety of the cofactor, involved in two H-bonding interactions with hinge residues Glu216 and Asp218.

We further evaluated the selectivity of the YpkA inhibitors by testing against two other kinases, MAPK and protein kinase C (PKC). It is not surprising that some compounds showed comparable inhibitory activities to MAPK, from which the homology models of YpkA were derived. For example, **1** showed the best inhibition of YpkA with an IC₅₀ of 1.81 μM but also exhibited similar activity to MAPK with an IC₅₀ of 2.45 μM. However, **2–4** are more selective to YpkA over MAPK and PKC with 5- to 10-fold better inhibition (Figure 4). The predicted interactions of **4** in the YpkA active site showed that the nitro group forms strong interactions with residue Arg221 at the end of hinge loop (Figure 5). Because an Asp or Glu residue is typically present at this position in mammalian serine/threonine kinases, interactions between the basic, positive charged Arg residue, and **4** may impart selectivity for YpkA. To the best of our knowledge, these are the first reported small-molecule inhibitors for YpkA, providing a means to investigate the mechanism of YpkA in bacterial pathogenesis, as well as a starting point for the design of potent and selective inhibitors as antiplague drugs. Although these results are

promising, one must consider nonspecific inhibition due to the induction of protein aggregation.²² On the basis of our experimental results in Figure 4, the reported inhibitors are most likely not acting as aggregation agents and are specifically inhibiting YpkA. This is demonstrated because the compounds have little effect on PKC but are inhibitory to MAPK, from which our homology model was created.

In summary, we have described an integrated approach combining machine learning techniques and high-throughput docking for the discovery of *Yersinia* protein kinase A inhibitors. We have made use of the abundant resource of known kinase inhibitors and have developed a SVM model to prioritize these databases. With the construction of homology models and an ensemble of protein structures, we performed multiple conformational high-throughput docking on the target-focused library for the search of potent and selective inhibitors of YpkA. The combination of both ligand-based and structure-based knowledge of protein kinases has demonstrated high screening efficiency and reasonable speed, which has allowed us to characterize the first reported inhibitors of *Yersinia* protein kinase A. This integrated approach therefore provides a practical method to account for protein flexibility in a large-scale database for virtual screening of effective inhibitors of therapeutic targets.

Acknowledgment. This work was funded in part by research funds to C.E.S. from the Rockefeller University and PHS Grant 1U19AI056510 from the National Institute of Allergy and Infectious Diseases.

Supporting Information Available: The 3D structural models of YpkA, detailed methods on homology modeling, SVM clustering, MD simulations, virtual screening, and YpkA purification, inhibitor data, and experimental details for Figure 4. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Galan, J. E.; Collmer, A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **1999**, *284*, 1322–1328.
- Stebbins, C. E. Structural insights into bacterial modulation of the host cytoskeleton. *Curr. Opin. Struct. Biol.* **2004**, *14*, 731–740.
- Henderson, D. A. The looming threat of bioterrorism. *Science* **1999**, *283*, 1279–1282.
- Marra, A. Targeting virulence for antibacterial chemotherapy: identifying and characterising virulence factors for lead discovery. *Drugs R&D* **2006**, *7*, 1–16.
- Perry, R. D.; Fetherston, J. D. *Yersinia pestis*—etiologic agent of plague. *Clin. Microbiol. Rev.* **1997**, *10*, 35–66.
- Letzelter, M.; Sorg, I.; Mota, L. J.; Meyer, S.; Stalder, J.; et al. The discovery of SycO highlights a new function for type III secretion effector chaperones. *EMBO J.* **2006**, *25*, 3223–3233.
- Juris, S. J.; Rudolph, A. E.; Huddler, D.; Orth, K.; Dixon, J. E. A distinctive role for the *Yersinia* protein kinase: actin binding, kinase activation, and cytoskeleton disruption. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9431–9436.
- Prehna, G.; Ivanov, M. I.; Bliska, J. B.; Stebbins, C. E. *Yersinia* virulence depends on mimicry of host rho-family nucleotide dissociation inhibitors. *Cell* **2006**, *126*, 869–880.
- Wiley, D. J.; Nordfeldth, R.; Rosenzweig, J.; DaFonseca, C. J.; Gustin, R.; et al. The Ser/Thr kinase activity of the *Yersinia* protein kinase A (YpkA) is necessary for full virulence in the mouse, mollifying phagocytes, and disrupting the eukaryotic cytoskeleton. *Microb. Pathog.* **2006**, *40*, 234–243.
- Navarro, L.; Koller, A.; Nordfeldth, R.; Wolf-Watz, H.; Taylor, S.; et al. Identification of a molecular target for the *Yersinia* protein kinase A. *Mol. Cell* **2007**, *26*, 465–477.
- Noble, M. E.; Endicott, J. A.; Johnson, L. N. Protein kinase inhibitors: insights into drug design from structure. *Science* **2004**, *303*, 1800–1805.
- Scapin, G. Protein kinase inhibition: different approaches to selective inhibitor design. *Curr. Drug Targets* **2006**, *7*, 1443–1454.
- Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209–225.

- (14) Muegge, I.; Enyedy, I. J. Virtual screening for kinase targets. *Curr. Med. Chem.* **2004**, *11*, 693–707.
- (15) Briem, H.; Gunther, J. Classifying “kinase inhibitor-likeness” by using machine-learning methods. *ChemBioChem* **2005**, *6*, 558–566.
- (16) Ford, M. G.; Pitt, W. R.; Whitley, D. C. Selecting compounds for focused screening using linear discriminant analysis and artificial neural networks. *J. Mol. Graphics Modell.* **2004**, *22*, 467–472.
- (17) Nolen, B.; Taylor, S.; Ghosh, G. Regulation of protein kinases; controlling activity through activation segment conformation. *Mol. Cell* **2004**, *15*, 661–675.
- (18) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377–395.
- (19) Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided. Mol. Des.* **2002**, *16*, 11–26.
- (20) Hardcastle, I. R.; Golding, B. T.; Griffin, R. J. Designing inhibitors of cyclin-dependent kinases. *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 325–348.
- (21) De Moliner, E.; Moro, S.; Sarno, S.; Zagotto, G.; Zanotti, G.; et al. Inhibition of protein kinase CK2 by anthraquinone-related compounds. A structural insight. *J. Biol. Chem.* **2003**, *278*, 1831–1836.
- (22) McGovern, S. L.; Shoichet, B. K. Kinase inhibitors: not just for kinases anymore. *J. Med. Chem.* **2003**, *46*, 1478–1483.

JM070645A