Ligand Design

Using Ligand-Mapping Simulations to Design a Ligand Selectively Targeting a Cryptic Surface Pocket of Polo-Like Kinase 1**

Yaw Sing Tan, Paweł Śledź, Steffen Lang, Christopher J. Stubbs, David R. Spring, Chris Abell,* and Robert B. Best*

The treatment of protein flexibility is a major challenge in structure-based drug design (SBDD)[1,2] as proteins are dynamic and commonly undergo conformational changes to bind ligands.[3–6] Consequently, binding sites may not be apparent in experimental structures of the unliganded protein. As a prototypical example, we focus here on the polo-box domain (PBD) of polo-like kinase 1 (Plk1), a serine/threonine kinase that is overexpressed in a wide range of cancers,[7,8] and is a known anticancer target due to its critical role in mitotic progression.[9] The PBD helps in subcellular localization of the protein by binding to serine- or threonine-phosphorylated sequences at a polar phosphopeptide binding site (Figure 1a).[10,11]

Recently, a secondary hydrophobic binding site (Figure 1b) proximal to the primary phosphopeptide binding site has been identified by two independent approaches.[12,13] Crystal structures have revealed that this pocket can accommodate hydrophobic side-chains of several ligands in slightly different binding modes, with a resulting increase in affinity;[12–16] however, when ligands are not bound to it, the pocket is closed. This cryptic pocket therefore presents a classic problem in SBDD targeting a flexible protein surface. Here, we show that although opening of the pocket is highly unfavorable in the absence of a ligand, it is possible to identify all of its known ligand-binding modes, as well as a novel mode, by using a modified ligand-mapping technique. The previously unknown binding mode was used as a basis for the design of a new ligand with similar affinity to others binding this pocket. The predictions were validated by solving the crystal structure of the bound complex.
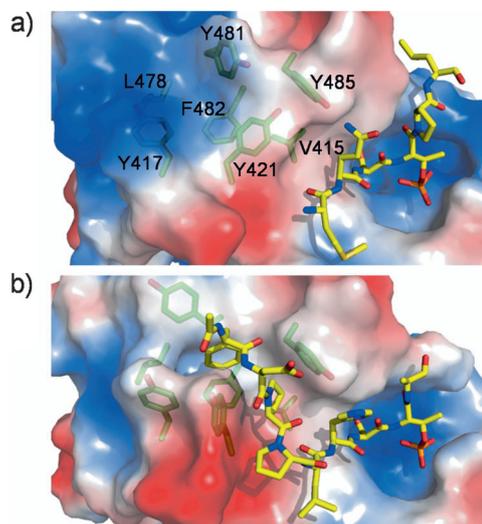
[*] Y. S. Tan,[+] Dr. P. Śledź,[+] Dr. S. Lang, C. J. Stubbs, Dr. D. R. Spring, Prof. C. Abell, Dr. R. B. Best
Department of Chemistry, University of Cambridge
Lensfield Road, Cambridge, CB2 1EW (UK)
E-mail: ca26@cam.ac.uk
    rbb24@cam.ac.uk

Dr. P. Śledź[+]
Present address: Department of Molecular Structural Biology
Max Planck Institute of Biochemistry
Am Klopferspitz 18, 82152 Martinsried (Germany)

[+] These authors contributed equally to this work.

Figure 1. Binding pockets of Plk1 PBD as revealed by phosphopeptide ligands. Residues of the hydrophobic binding site are shown in green sticks and ligands in yellow sticks. Regions of negative, positive and neutral electrostatic potential on the protein surface are red, blue and white, respectively. a) Crystal structure of a phosphopeptide bound to the positively charged phosphopeptide binding pocket (PDB code 1Q4K), with closed hydrophobic pocket. b) Crystal structure of a phosphopeptide bound to both binding pockets (PDB code 3P37), with open hydrophobic pocket.

The walls of the hydrophobic PBD secondary binding site comprise Tyr417, Tyr421, Leu478, Tyr481 and Tyr485, with Val415 and Phe482 lying at the bottom. Side-chain movements of Tyr417 and Tyr481 allow for the accommodation of a phenyl or other hydrophobic moiety. We first attempted to generate a conformational ensemble of the hydrophobic binding site by performing a 50 ns molecular dynamics (MD) simulation of the unliganded protein in explicit water, using the Amber ff99SB-ILDN force field.[17,18] This approach is similar to the relaxed complex scheme successfully used in other studies.[19–21] Root-mean-square deviation (RMSD)-based clustering of the MD trajectory was performed to compare the conformations of the binding site with those seen before in various crystal structures. Four distinct closed conformations of the site were identified, three of which are also observed in crystal structures (Figure 2a–d).

In contrast, no conformations in which the key residue Tyr481 is "open" (e.g. Figure 2e,f) were seen during the simulation. The high propensity of Tyr481 to adopt its closed conformation was confirmed by a second simulation of the unliganded PBD, starting from a crystal structure with the pocket open, after removing the ligand (PDB code 3P37).
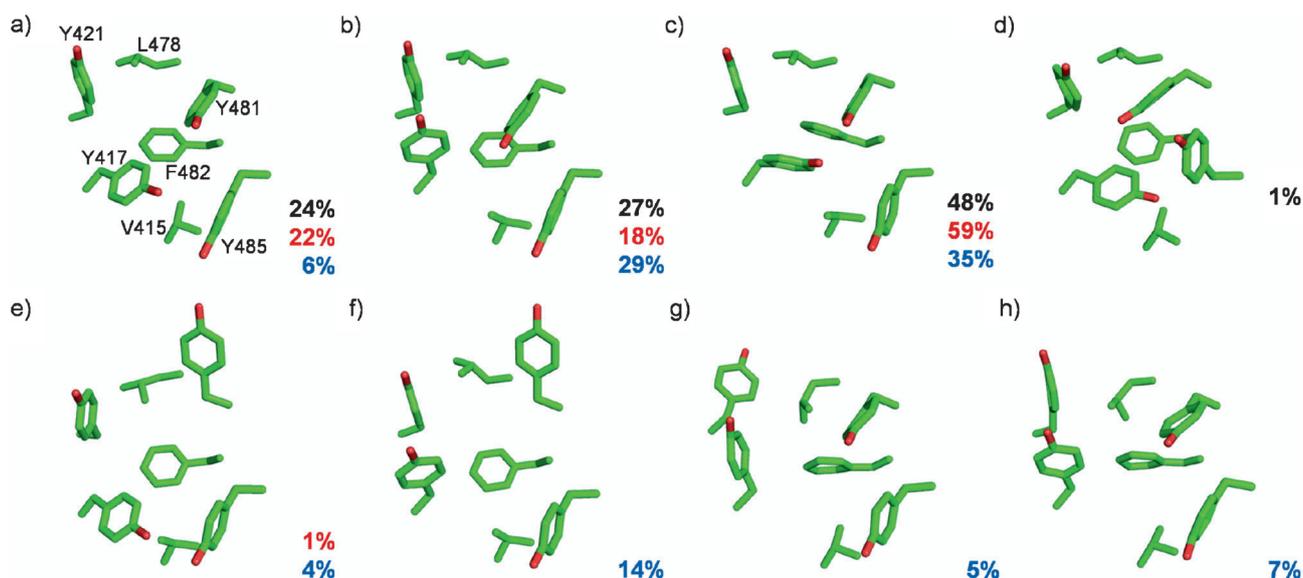
Figure 2. Conformations of the hydrophobic binding pocket in Plk1 PBD. Percentage populations of the conformations observed in the single long, multiple short and ligand-mapping simulations are indicated in black, red and blue, respectively. a–c) Closed conformations observed in crystal structures. d) Closed conformation observed only in the single long MD simulation. e,f) Open conformations observed in crystal structures. g,h) New conformations observed in ligand-mapping MD simulations.

Tyr481 closed over the pocket after about 2 ns and remained so for the rest of the 50 ns simulation (see Supporting Information, Figure S1b). In an attempt to maximize the coverage of conformational space, we also ran 10 independent 5 ns simulations instead of a single long one.[22] Opening of Tyr481 was observed in one of these runs, however the pocket did not fully open as Tyr417 remained in its closed conformation (Figure 2e). When this simulation was extended, Tyr481 closed after 2 ns and did not reopen.

To gain more insight into the lack of pocket opening events, we estimated the free energy for opening by using umbrella sampling[23] to bias the side-chain of Tyr481 to sample different $\chi 1$ torsion angle values. The free energy difference between its open and closed conformations was estimated to be ca. 4 kcal mol^{-1} with a barrier of ca. 6 kcal mol^{-1} (Figure S2). This explains the infrequent sampling of open conformations by Tyr481 within the timescale of the unliganded simulations, and its rapid closure in the absence of a stabilizing ligand.

These observations highlight the difficulties of using unbiased MD trajectories to identify cryptic binding sites when ligand-binding protein conformations are rarely sampled in the absence of a ligand: "solvent mapping" approaches, in which binding sites are identified by scanning model ligands over the protein surface,[24] require the sites to be accessible. Methods such as replica exchange MD[25] and metadynamics[26] can improve sampling, but are not specific to opening hydrophobic pockets. Here, we adopt an alternative enhanced sampling scheme, suggested by the open conformations in crystal structures all being stabilized by either a hydrophobic ligand or crystal contact.[12] We thus increased the sampling of relevant ligand-binding conformations by incorporating benzene molecules into the simulations, due to the pocket's known affinity for a phenyl moiety.[12,13]

Although hydrophilic molecules have previously been used for this purpose at high (2.7 M) concentrations,[27] a lower concentration of 0.2 M was needed in this work to avoid phase separation of benzene and water. Note that an alternative method, site-identification by ligand competitive saturation (SILCS),[28–30] uses high concentrations of hydrophobic ligands (1 M) in conjunction with an inter-ligand repulsive interaction energy term. However, SILCS's main motivation is to characterize the protein surface through *competition* between ligands representative of different functionalities. Our aim is to identify only binding pockets with a high affinity for aromatic groups, in which case a lower benzene concentration is more selective. The absence of a repulsive potential also allows for accommodation of multiple benzene molecules in the same binding pocket—as observed in some of our simulations.

For optimal conformational sampling, 10 independent 5 ns ligand-mapping simulations were performed with different initial benzene distributions. All binding pocket conformations previously seen in crystal structures were reproduced by the simulations (Figure 2). The open pocket conformations were stabilized by the binding of benzene (Figure 3 a,b). Mapping of benzene occupancy for trajectory structures with similar benzene binding modes at the pocket correlated well with previously seen interactors of the pocket, showing that our ligand-mapping simulations reproduce crystallographic binding modes of ligands on the PBD surface (Figure 3 c,d). In addition, in some of the simulations, we observed "half-open" conformations of the pocket (Figure 2 g,h) not seen in crystal structures. These conformations are involved in a new "half open" binding mode in which a benzene molecule lodges in a smaller cleft formed by five of the hydrophobic pocket residues: Val415, Tyr417, Tyr481, Phe482 and Tyr485 (Figure 4 a).
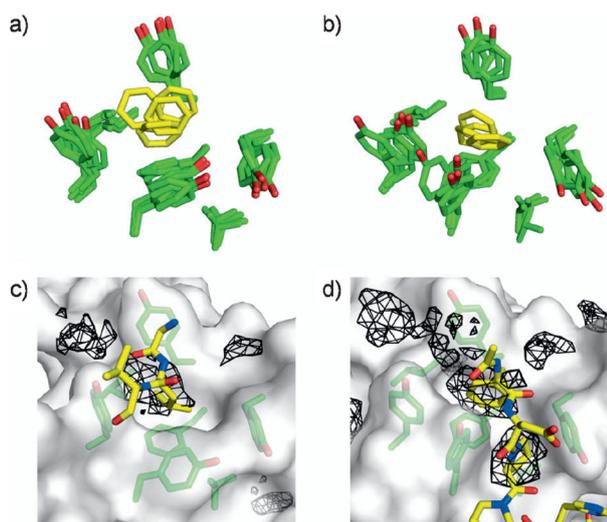
Figure 3. Reproduction of crystallographic binding modes at PBD hydrophobic pocket by ligand-mapping simulations. a,b) Superposition of five snapshots of the partially and fully open binding pocket, respectively, from ligand-mapping simulations. Pocket residues are green and benzene molecules are yellow. c) Benzene probes mimic the crystal-packing interaction of Leu394 from a neighboring protomer with the binding pocket (PDB code 3P35). Interacting ligands and crystal contact residues are shown in yellow sticks while regions sampled by benzene probes for this protein conformation are represented as black mesh (see Supporting Information for details). Only the part of the protomer binding to the pocket is shown. d) Benzene probes mimic the interactions of Phe and Pro from the peptide ligand with the binding pocket in the PBD/FDPPLHSpTA complex (PDB code 3P37).
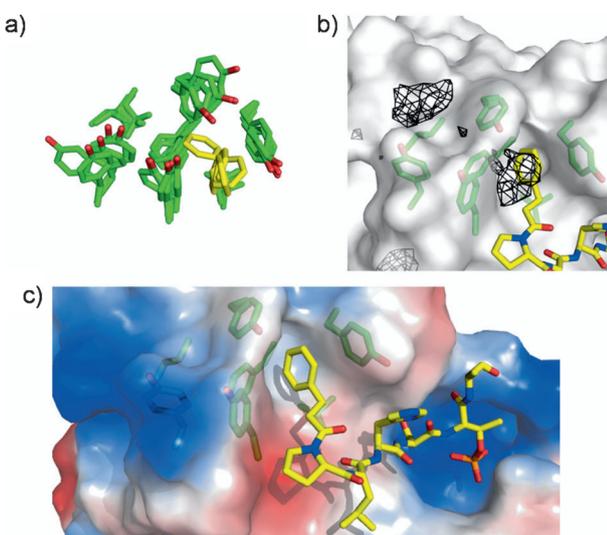

Figure 4. Design of a chimeric peptide ligand based on an alternative binding mode obtained from ligand-mapping simulations. a) Superposition of five snapshots of PBD binding pocket (green) from a ligand-mapping MD trajectory, showing an alternative binding mode of benzene (yellow). b) Binding mode of chimeric peptide (yellow sticks) at PBD pocket (green sticks) overlaid with benzene probability isosurfaces (black mesh) for the protein structure shown. c) Crystal structure of PBD complexed with chimeric peptide (yellow sticks). Regions of positive and negative electrostatic potential on the PBD surface are colored blue and red, respectively.

The generation of previously observed binding modes at the pocket in the ligand-mapping simulations is encouraging. To further validate the simulation results, we designed a ligand to present a phenyl ring to the pocket in the half-open conformation seen in the simulations (Figure 4a). Based on the known structure of FDPPLHSpTA bound to PBD (Figure 1b), the new ligand was modeled by replacing the N-terminal FDP residues with a 3-phenylpropanoyl moiety, which incorporates a minimal aliphatic linker between the phenyl group in the pocket and the truncated peptide. Evidence for the stability of this chimeric peptide ligand at the pocket was provided by an MD simulation of its complex with PBD (Figure S3, S4 and Table S1).

We synthesized the ligand and determined its K_D to be 330 nM by isothermal titration calorimetry (ITC), comparable to that of the significantly larger FDPPLHSpTA peptide (250 nM). A crystal structure of the complex (PDB code 4E67) was also obtained, showing that the binding mode of the new chimeric ligand at the chosen binding pocket is indeed similar to that envisaged in the design (Figure 4b,c). This demonstrates specificity in our design, as other phosphopeptides are known to bind in a completely different mode, avoiding the pocket altogether.[12]

We have shown that the introduction of benzene probe molecules into MD simulations considerably expands the accessible conformational space of the flexible PBD hydrophobic pocket. We have also used this ligand-mapping approach for the first time to design a novel high-affinity ligand, whose binding mode was confirmed by X-ray crystallography. Simulations with explicit solvent are undoubtedly expensive for drug design. However, our results provide a proof of principle which may be used as the basis for more efficient sampling strategies in future, for example by using probe molecules in conjunction with an implicit water model.[31,32] In this work, we have focused on a hydrophobic pocket. However, there is no reason that a library of small molecules presenting different chemical functionalities could not be similarly used to expose other types of cryptic binding sites.

[1] P. Cozzini, G. E. Kellogg, F. Spyrakis, D. J. Abraham, G. Costantino, A. Emerson, F. Fanelli, H. Gohlke, L. A. Kuhn, G. M. Morris, M. Orozco, T. A. Pertinhez, M. Rizzi, C. A. Sotriffer, *J. Med. Chem.* **2008**, *51*, 6237–6255.
[2] G. Fuentes, S. G. Dastidar, A. Madhumalar, C. S. Verma, *Drug Dev. Res.* **2011**, *72*, 26–35.
[3] M. J. Betts, M. J. E. Sternberg, *Protein Eng.* **1999**, *12*, 271–283.
[4] I. M. Withers, M. P. Mazanetz, H. Wang, P. M. Fischer, C. A. Laughton, *J. Chem. Inf. Model.* **2008**, *48*, 1448–1454.
[5] W. L. DeLano, M. H. Ultsch, A. M. de Vos, J. A. Wells, *Science* **2000**, *287*, 1279–1283.
[6] S. J. Teague, *Nat. Rev. Drug Discovery* **2003**, *2*, 527–541.

[7] U. Holtrich, G. Wolf, A. Brauninger, T. Karn, B. Bohme, H. Rubsamenwaigmann, K. Strebhardt, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1736–1740.

[8] D. M. Lowery, D. Lim, M. B. Yaffe, *Oncogene* **2005**, *24*, 248–259.

[9] K. Strebhardt, A. Ullrich, *Nat. Rev. Cancer* **2006**, *6*, 321–330.

[10] S. Song, T. Z. Grenfell, S. Garfield, R. L. Erikson, K. S. Lee, *Mol. Cell. Biol.* **2000**, *20*, 286–298.

[11] N. Reynolds, H. Ohkura, *J. Cell Sci.* **2003**, *116*, 1377–1387.

[12] P. Śledź, C. J. Stubbs, S. Lang, Y. Q. Yang, G. J. McKenzie, A. R. Venkitaraman, M. Hyvonen, C. Abell, *Angew. Chem.* **2011**, *123*, 4089–4092; *Angew. Chem. Int. Ed.* **2011**, *50*, 4003–4006.

[13] F. Liu, J.-E. Park, W.-J. Qian, D. Lim, M. Graeber, T. Berg, M. B. Yaffe, K. S. Lee, T. R. Burke, Jr., *Nat. Chem. Biol.* **2011**, *7*, 595–601.

[14] F. Liu, J.-E. Park, W.-J. Qian, D. Lim, A. Scharow, T. Berg, M. B. Yaffe, K. S. Lee, T. R. Burke, Jr., *ACS Chem. Biol.* **2012**, *7*, 805–810.

[15] F. Liu, J.-E. Park, W.-J. Qian, D. Lim, A. Scharow, T. Berg, M. B. Yaffe, K. S. Lee, T. R. Burke, Jr., *ChemBioChem* **2012**, *13*, 1291–1296.

[16] P. Śledź, S. Lang, C. J. Stubbs, C. Abell, *Angew. Chem.* **2012**, *124*, 7800–7803; *Angew. Chem. Int. Ed.* **2012**, *51*, 7680–7683.

[17] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins Struct. Funct. Bioinf.* **2006**, *65*, 712–725.

[18] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins Struct. Funct. Bioinf.* **2010**, *78*, 1950–1958.

[19] J. H. Lin, A. L. Perryman, J. R. Schames, J. A. McCammon, *J. Am. Chem. Soc.* **2002**, *124*, 5632–5633.

[20] J. R. Schames, R. H. Henchman, J. S. Siegel, C. A. Sotriffer, H. H. Ni, J. A. McCammon, *J. Med. Chem.* **2004**, *47*, 1879–1881.

[21] L. S. Cheng, R. E. Amaro, D. Xu, W. W. Li, P. W. Arzberger, J. A. McCammon, *J. Med. Chem.* **2008**, *51*, 3878–3894.

[22] L. S. D. Caves, J. D. Evanseck, M. Karplus, *Protein Sci.* **1998**, *7*, 649–666.

[23] G. M. Torrie, J. P. Valleau, *J. Comput. Phys.* **1977**, *23*, 187–199.

[24] A. Miranker, M. Karplus, *Proteins Struct. Funct. Genet.* **1991**, *11*, 29–34.

[25] Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **1999**, *314*, 141–151.

[26] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12566.

[27] J. Seco, F. J. Luque, X. Barril, *J. Med. Chem.* **2009**, *52*, 2363–2371.

[28] O. Guvench, A. D. MacKerell, Jr., *PLoS Comp. Biol.* **2009**, *5*, e1000435.

[29] E. P. Raman, W. Yu, O. Guvench, A. D. MacKerell, Jr., *J. Chem. Inf. Model.* **2011**, *51*, 877–896.

[30] T. J. Foster, A. D. Mackerell, Jr., O. Guvench, *J. Comput. Chem.* **2012**, *33*, 1880–1891.

[31] U. Haberthür, A. Caflisch, *J. Comput. Chem.* **2008**, *29*, 701–715.

[32] V. Zoete, A. Grosdidier, M. Cuendet, O. Michielin, *J. Mol. Recognit.* **2010**, *23*, 457–461.