Subpocket-based fingerprint for structural kinase comparison

Dominique Sydow[1], Eva Aßmann[1], Albert Kooistra[2], Friedrich Rippmann[3], Andrea Volkamer[1]

[1] *In silico* Toxicology, Institute for Physiology, Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany

[2] Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Jagtvej 162, DK-2100 Copenhagen, Denmark

[3] Computational Chemistry & Biology, Merck Healthcare KGaA, Frankfurter Str. 250, 64293 Darmstadt, Germany

Introduction

Kinases are important and well studied drug targets for cancer and inflammatory diseases. Due to the highly conserved structure of kinases, especially at the ATP binding site, the main challenge when developing kinase inhibitors is achieving selectivity, which requires a comprehensive understanding of kinase similarity. [1] Here, we present our work on a novel fingerprinting strategy designed specifically for kinase pockets, allowing for similarity studies across the structurally covered kinome.

Methods

The kinase fingerprint is based on the KLIFS [2] pocket alignment, which defines 85 aligned pocket residues for all kinase structures. This enables a residue-by-residue comparison across the kinome without a computationally expensive alignment step.

Data preparation. KLIFS dataset was filtered for entries (i) describing human kinases, (ii) in DFG-in conformation, (iii) with the best quality score per PDB structure (if multiple chains and/or alternate models available), and (iv) with a resolution ≤ 4 Å and a quality score ≥ 4. The resulting kinase dataset consists of 3,875 structures, representing 253 kinases.

Kinase fingerprint. The *pocket fingerprint* consists of 85 concatenated *residue fingerprints*, each encoding a residue's spatial and physicochemical properties (**Fig. 1a** and **2a**). The *physicochemical properties* encompass for each residue its size, side chain orientation and pharmacophoric features as described by SiteAlign [3], in addition to its solvent exposure as implemented in Biopython's module `Bio.PDB.HSExposure` [4, 5]. Inspired by the ligand-based USR approach [6], the *spatial properties* describe the residue's position in relation to the kinase pocket centroid and important kinase subpockets, i.e. the hinge region, the DFG region, and the front pocket (**Fig. 1b**). The resulting distance distributions per subpocket are reduced in complexity to the first three moments, i.e. the mean, variance, and skewness.

Kinase comparison & scoring. Kinase structures are compared pairwise using the inverse, translated and scaled Manhattan distance as implemented for the USR method [3]. For each kinase pair, the best scoring structure pair is used for further analysis, resulting in a 253x253 similarity matrix.

$$d(A,B) = \frac{1}{1 + \frac{1}{N}\sum_{i=1}^{N} |A_i - B_i|} \in]0,1]$$

inverse; translated; scaled; Manhattan distance

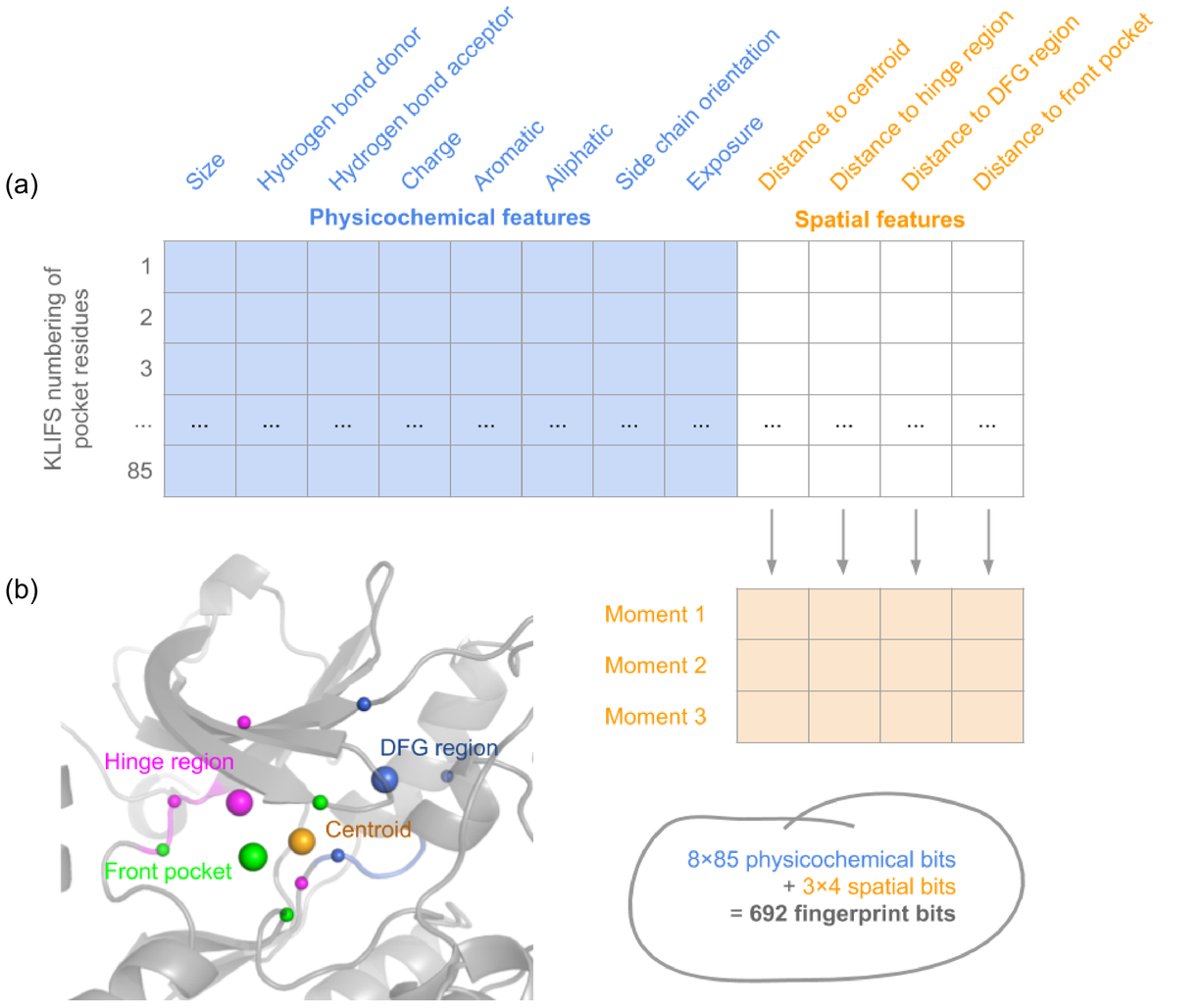

Fig. 1: **(a)** Composition of the 692 bit kinase fingerprint. **(b)** Reference points for spatial feature calculation: centroid (orange), hinge region (magenta), DFG region (blue), and front pocket (green). The last three reference points (large spheres) are the centroids of three anchor residue CA atoms each (small spheres). Backbones of hinge and DFG region are highlighted.

Results

The potential of our subpocket-based kinase comparison is demonstrated by uncovering retrospectively on- and off-targets for EGFR inhibitor erlotinib using KinMap [7] (**Fig. 2b**). Compared to erlotinib profiling data by Karaman et al. [8], our 20 most similar structures to EGFR include many off-targets in the TK group as well as the reported main off-target kinases LOK and SLK (in the more distant STE group), though missing the off-target kinase GAK (Other).

Furthermore, clustering of all similarity scores for the kinase pairs (253x253 similarity matrix) shows that our method can reproduce the Manning classification [9] in large part, however also reveals new relationships such as the aforementioned STE kinases within the TK group or the grouping of DRAK2 (DAPK) with CaMKK2 (Other), a reported off-target of DRAK2 inhibitors [10] (**Fig. 2c**).

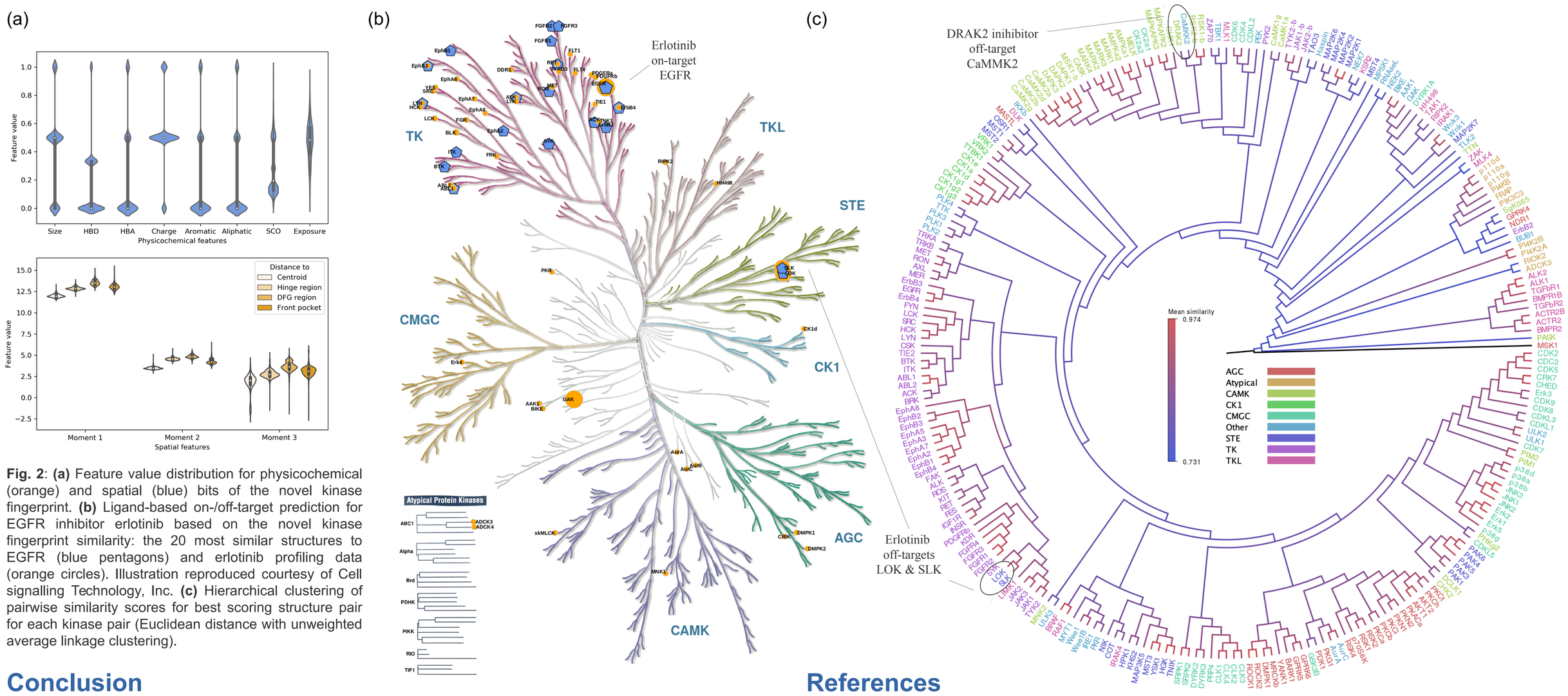

Fig. 2: **(a)** Feature value distribution for physicochemical (orange) and spatial (blue) bits of the novel kinase fingerprint. **(b)** Ligand-based on-/off-target prediction for EGFR inhibitor erlotinib based on the novel kinase fingerprint similarity: the 20 most similar structures to EGFR (blue pentagons) and erlotinib profiling data (orange circles). Illustration reproduced courtesy of Cell signalling Technology, Inc. **(c)** Hierarchical clustering of pairwise similarity scores for best scoring structure pair for each kinase pair (Euclidean distance with unweighted average linkage clustering).

Conclusion

Our subpocket-based kinase fingerprinting strategy can partially retrieve the Manning kinase classification but also reveals structural relationships between different kinase groups. Therefore, we believe our fingerprint can help researchers (i) to detect potential promiscuities and off-targets at an early stage of inhibitor design and (ii) to conduct structure-informed polypharmacology studies.

References

[1] Kooistra and Volkamer. Ann Rep Med Chem. 2017, 50, 263-299. [2] van Linden et al. J Med Chem. 2014, 57, 249-77. [3] Schalon et al. Proteins. 2008, 71, 1755-78. [4] Cock et al. Bioinformatics 2009, 25, 1422-3. [5] Hamelryck. Proteins 2005, 59, 38-48. [6] Ballester and Richards. J Comp Chem. 2007, 28, 1711-23. [7] Eid et al. BMC Bioinf 2017, 18. [8] Karaman et al. Nature Biotech 2008, 26, 127-32. [9] Manning et al. Science. 2002, 298, 1912-34. [10] Picado et al. ACS National Meeting Orlando 2019 (poster).