ELSEVIER

Computational functional group mapping for drug discovery

Olgun Guvench

SilcsBio, LLC, 8 Market Street, Suite 300, Baltimore, MD 21202, USA

Computational functional group mapping (cFGM) is emerging as a high-impact complement to existing widely used experimental and computational structure-based drug discovery methods. cFGM provides comprehensive atomic-resolution 3D maps of the affinity of functional groups that can constitute drug-like molecules for a given target, typically a protein. These 3D maps can be intuitively and interactively visualized by medicinal chemists to rapidly design synthetically accessible ligands. Given that the maps can inform selection of functional groups for affinity, specificity, and pharmacokinetic properties, they are of utility for both the optimization of existing drug candidates and creating novel ones. Here, I review recent advances in cFGM with emphasis on the unique information content in the approach that offers the potential of broadly facilitating structure-based ligand design.

Introduction

The design of small molecules that bind with optimal specificity and affinity to their biological targets, typically proteins, is based on the idea of complementarity between the functional groups in a small molecule and the binding site of the target. The complementarity arises from fundamental physical interactions, such as hydrogen bonds, salt bridges, and hydrophobic association. This basic intuitive idea suggests another one: 'functional group mapping' (FGM), which is simply to make a map, for each functional group, of the regions in the vicinity of the target surface with which that functional group preferentially binds. With such a map in hand for each functional group in their synthetic palette, medicinal chemists can focus their efforts on designing small molecules that best match the maps. Ideally, FGM should consider not only the target surface as seen in an atomic-resolution crystal structure, but also inducible pockets resulting from the combination of target flexibility and functional group binding [1]. With such complete information, the medicinal chemist could design the most efficacious small molecule for a binding site on the target protein.

Experimental FGM

There have been numerous experimental and computational approaches to FGM. One approach that can be considered FGM

is the synthesis and testing of a series of similar drug-like molecules. This approach, resulting in a structure–activity relation (SAR), can reveal which functional groups have beneficial interactions with the target. The SAR approach can be extended to include 3D information about the drug-like molecules in the series in a quantitative manner (3D-QSAR) [2,3]. Based on the assumption that the molecules bind in the same pose, one can deduce which functional groups at what locations interact best with what parts of the target binding site, and this approach can be particularly powerful given a known binding pose in the context of the target [4]. The results can be used as a map to guide the creation of new compounds with improved efficacy or to scaffold-hop to other compounds with similar 3D placement of the relevant functional groups [5–7]. Another FGM approach involves testing the ability of small, simple compounds (molecular weight 100–300 Da) to bind to a target. These compounds are sufficiently small as to have limited internal degrees of conformational freedom, and often correspond to a single functional group that might occur in a drug-like molecule making them, in a sense, 'fragments' of drug-like molecules. Fragments can be screened in high-concentration bioassays with nuclear magnetic resonance (NMR)-based hit confirmation followed by X-ray structure determination of target–fragment complexes to determine the binding poses of hits [8–12]. Using the known binding poses of several fragments, medicinal chemists can synthetically link fragments together into a drug-like

E-mail address: oguvench@silcsbio.com.

molecule and then test this molecule for selectivity and affinity, a process commonly referred to as 'fragment-based drug design' (FBDD) [13–15].

Computational FGM

The above SAR-based approaches, by their very design, combine experimental data with computational modeling. By contrast, target-based computational approaches have evolved in parallel with experimental fragment-based FGM. The earliest of these purely cFGM approaches date back to the earliest days of computer-aided drug design (CADD) and, therefore, the following is necessarily only a partial list: (i) GRID, where the interaction energy between a probe group (e.g., methyl) and the target is computed on a grid encompassing the binding site to give an array of energy values that can be displayed as contour surfaces [16]; (ii) MCSS, where multiple non-interacting fragments are energy minimized or quenched in the presence of the target to determine optimal binding poses [17]; (iii) FTMAP, which combines a grid approach with determination of low-energy conformations of fragments and clustering to determine favorable binding sites [18,19]; (iv) alchemical transformation, where the free energy of morphing one functional group into another is computed, either in a discrete (e.g., free energy perturbation or thermodynamic integration) or continuous (e.g., lambda dynamics) manner [20–22]; (v) WaterMap, which determines the locations and thermodynamic properties of water molecules that solvate protein-binding sites [23]; and (vi) ligand or co-solvent mapping simulations, which determine the locations of fragment molecules that can occupy protein-binding sites [24–28].

The benefits of using computational approaches are well known: they have the potential to save time, materials, and labor. Their main liability is also well known: the reliability of the results depends on the accuracy of the computational model and the need for adequate conformational sampling [29–32]. Therefore, a key driver of computational methods development is to retain the benefits while increasing the reliability. Much of these developments have been enabled by the steady improvement in computing hardware and algorithms [33,34] along with the maturation of all-atom explicit-solvent force fields [31,35,36]. These developments allow for the routine modeling of systems of relevant size, including explicit solvent, into the microsecond timescale using molecular dynamics (MD) simulation.

cFGM with all-atom explicit-solvent MD

The application of all-atom explicit-solvent MD brings two key improvements in reliability to cFGM. The first is the discrete nature of molecular water, which allows for water to explicitly compete with functional groups, to act as a bridging molecule in enclosed spaces, and to contribute to the hydrophobic effect. The second is that target flexibility is naturally incorporated as the system samples thermally accessible conformations and, therefore, inducible pockets that are not evident from analysis of crystallographic structures. Thus, two key challenges in cFGM and in CADD in general (solvation effects and target flexibility) are accounted for in a physically rigorous fashion. Reflecting this, of the cFGM approaches listed above, both (v) and (vi) use the all-atom explicit-solvent MD framework, and

(iv) is increasingly used in this framework. The remainder of this brief review focuses on approach (vi), because it is the closest analog to experimental fragment-based FGM and has seen many developments in recent years. cFGM involving all-atom explicit-solvent MD includes co-solvent [24], 'MixMD' [27], Site-Identification by Ligand Competitive Saturation ('SILCS') [25], and related methods [26,28]. Using a single solute, such as isopropanol [24], acetonitrile [27], or chlorobenzene [37], in an aqueous solution provides functional group maps for that particular solute. It is also possible to use multiple fragments in the same simulation, each representing different functional groups [25,28,38]. This allows one to generate maps for, for example, hydrogen bond donors, hydrogen bond acceptors, aliphatic groups, aromatic groups, cations, and anions all from a single simulation. A more recent development is the application of sophisticated sampling methods to target buried or occluded pockets [39–41].

Advantages to cFGM with all-atom explicit-solvent MD

There are three scientific advantages to cFGM using all-atom explicit-solvent MD simulations over experimental methods. The first is the possibility of detecting low-affinity binding regions. Given that an MD trajectory captures not only long-lasting, but also fleeting interactions between fragments and the target, low-affinity binding regions, including their specific locations on the target, are detectable. The second advantage is an outcome from the first: the FGM for all functional groups are present for all regions of the entire target structure. For example, while a particular region might favor a hydrogen-bond donor, analysis of the FGMs might reveal that an aliphatic functional group can also favorably occupy that same region, information that is of high value to a medicinal chemist. A third is the ability to prevent aggregation of hydrophobic fragments and fragment-induced denaturation of the target. Ligand aggregation and target denaturation in experimental binding assays can lead to false positives or negatives [42,43] Several different approaches have been described for preventing target denaturation, including the use of weak restraining potentials on the target atoms (while still retaining e.g., full side-chain and loop flexibility) [25,44], limiting simulation timescales to below the kinetic limit for denaturation [27], and postsimulation analysis and exclusion of trajectory snapshots in which the target has denatured [44]. A novel approach to preventing aggregation has been described in the context of SILCS [25]. There, an additional interaction site is added to each fragment and serves to repel fragments that get within contact distance while all other interactions in the system are left unperturbed. Therefore, cFGM can be scientifically advantageous over experimental approaches when mapping denaturing or hydrophobic functional groups.

An additional benefit of cFGM in the all-atom explicit-solvent MD framework is the easy, intuitive way in which data can be visualized to guide small-molecule design. The simulation trajectory data can be converted to 3D maps, where binding probability p is determined as a function of x, y, z location for each point around the entire target structure. Using a grid resolution of ~1 Å results in fine-grained maps (see [45] for a sample workflow) that can be displayed using existing standard file formats (e.g., CCP4 [46], as used for crystallographic electron densities, or AutoDock
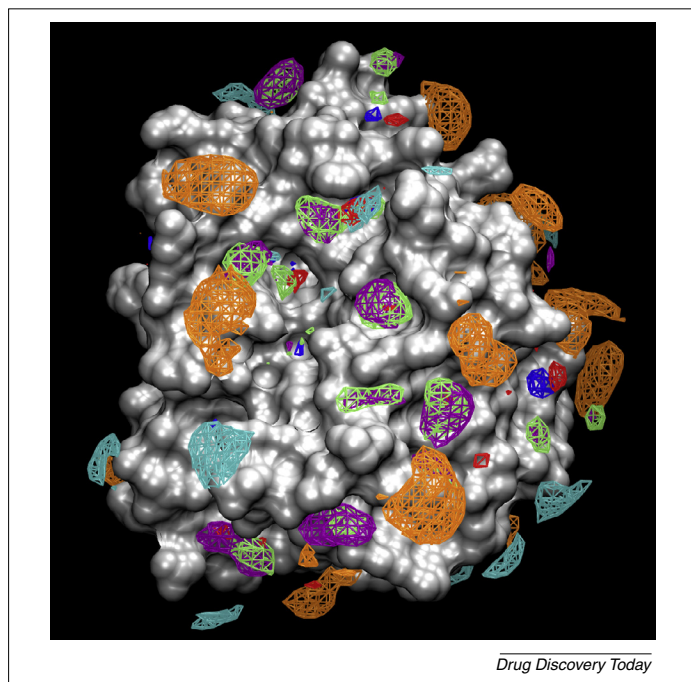
FIGURE 1

Computational functional group mapping (cFGM) of Factor Xa using the SILCS methodology (data available for free download at silcsbio.com). Maps are for benzene (purple), propane (green), neutral hydrogen bond donors (blue), neutral hydrogen bond acceptors (red), methylammonium (cyan), and acetate (orange).

FIGURE 2

Overlay of benzene functional group map from Fig. 1 (main text) with a known Factor Xa ligand in its crystallographic binding pose. Contours have been adjusted to show both (**a**) low-affinity and (**b**) high-affinity benzene binding. The solid-purple arrow shows an area that the mapping confirms is important for ligand binding, and the dashed-purple arrow shows an area for possible extension of the ligand through the addition of an aromatic functional group.

[47], as used for docking grids). Therefore, the 3D map for each functional group (i.e., the FGMs) can be loaded and viewed with the crystal structure of the target (Fig. 1). Given that standard molecular visualization programs permit interactive adjustment of the contour level for $p(x, y, z)$ data, the medicinal or computational chemist can adjust the contour level while rotating the system to see all regions with some binding, or regions with only high-affinity binding, or even unfavorable binding regions. Notably, the experimentally determined coordinates for a molecule bound to the target can be overlaid on the FGMs, because the FGMs themselves are ligand independent. The chemist can then zoom in on the area of interest and use the visual information in a qualitative manner to either plan extensions to a known molecule or design entirely new compounds (Fig. 2).

Quantitative application

In addition to the direct, interactive, intuitive qualitative application of 3D functional group maps, automated, quantitative applications also exist. One example is to use the maps to generate 3D pharmacophore models [48,49]. These can then be applied using existing standard virtual database screening tools to identify compounds. Another example is the direct use of the maps as scoring functions for *in silico* docking and pose refinement. Importantly, the scoring function can be expressed in units of free energy, ΔG, using the simple, theoretically exact relation $\Delta G = -RT \ln(p)$ [28,50]. This has the additional benefit of discriminating fragment poses that are most likely to be retained upon fragment elaboration into larger ligands because of the relation
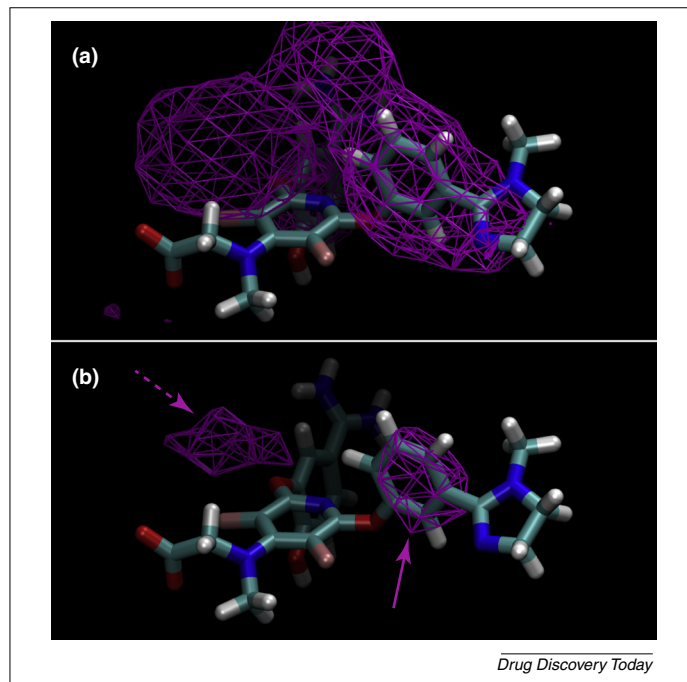
between fragment-binding free energy and conservation of binding pose [51]. Notably, because the FGMs are precomputed, docking and scoring can be performed in a rapid fashion as required for iterative ligand design. Alternatively, MD snapshots of target-fragment complexes can be used in a much more rapidly computed alternative to traditional alchemical transformation [(iv) above] [52].

Concluding remarks

All-atom explicit-solvent MD cFGM approaches are becoming increasingly better characterized and more applied [28,53–60]. This emerging cFGM methodology offers medicinal chemists an intuitively and visually accessible tool to facilitate ligand design, and the method can be extended to perform quantitative estimates of relative ligand affinity. Therefore, I expect cFGM to become a routine complement to the existing workflow of industrial medicinal and computational chemists.

Conflicts of interest

O.G. is a co-inventor of the SILCS cFGM technology, and is managing director and a co-owner of SilcsBio, LLC, which is commercializing this technology.

References

1 Johnson, D.K. and Karanicolas, J. (2013) Druggable protein interaction sites are more predisposed to surface pocket formation than the rest of the protein surface. *PLoS Comput. Biol.* 9, e1002951

2 Potemkin, V. and Grishina, M. (2008) Principles for 3 D/4 D QSAR classification of drugs. *Drug Discov. Today* 13, 952–959

3 Wang, T. *et al.* (2015) Quantitative structure–activity relationship: promising advances in drug discovery platforms. *Expert Opin. Drug Discov.* 10, 1283–1300

4 Fang, C. and Xiao, Z. (2015) Receptor-based 3 D-QSAR in drug design: methods and applications in kinase studies. *Curr. Top. Med. Chem.* 16, 1463–1477

5 Tresadern, G. and Bemporad, D. (2010) Modeling approaches for ligand-based 3 D similarity. *Future Med. Chem.* 2, 1547–1561

6 Verma, J. *et al.* (2010) 3D-QSAR in drug design – a review. *Curr. Top. Med. Chem.* 10, 95–115

7 Sun, H. *et al.* (2012) Classification of scaffold-hopping approaches. *Drug Discov. Today* 17, 310–324

8 Hesterkamp, T. and Whittaker, M. (2008) Fragment-based activity space: smaller is better. *Curr. Opin. Chem. Biol.* 12, 260–268

9 Davies, T.G. and Tickle, I.J. (2012) Fragment screening using X-ray crystallography. *Top. Curr. Chem.* 317, 33–59

10 Allen, K.N. *et al.* (1996) An experimental approach to mapping the binding surfaces of crystalline proteins. *J. Phys. Chem.* 100, 2605–2611

11 Mattos, C. and Ringe, D. (1996) Locating and characterizing binding sites on proteins. *Nat. Biotechnol.* 14, 595–599

12 Liepinsh, E. and Otting, G. (1997) Organic solvents identify specific ligand binding sites on protein surfaces. *Nat. Biotechnol.* 15, 264–268

13 Hennig, M. *et al.* (2012) Combining biophysical screening and X-ray crystallography for fragment-based drug discovery. *Top. Curr. Chem.* 317, 115–143

14 Erlanson, D.A. (2012) Introduction to fragment-based drug discovery. *Top. Curr. Chem.* 317, 1–32

15 Scott, D.E. *et al.* (2012) Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* 51, 4990–5003

16 Goodford, P.J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28, 849–857

17 Miranker, A. and Karplus, M. (1991) Functionality maps of binding-sites – a multiple copy simultaneous search method. *Proteins* 11, 29–34

18 Ngan, C.H. *et al.* (2012) FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res.* 40, W271–W275

19 Dennis, S. *et al.* (2002) Computational mapping identifies the binding sites of organic solvents on proteins. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4290–4295

20 Knight, J.L. and Brooks, C.L., 3rd (2009) Lambda-dynamics free energy simulation methods. *J. Comput. Chem* 30, 1692–1700

21 Reddy, M.R. *et al.* (2014) Free energy calculations to estimate ligand-binding affinities in structure-based drug design. *Curr. Pharm. Des.* 20, 3323–3337

22 Gilson, M.K. and Zhou, H.X. (2007) Calculation of protein–ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* 36, 21–42

23 Young, T. *et al.* (2007) Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proc. Natl. Acad. Sci. U. S. A.* 104, 808–813

24 Seco, J. *et al.* (2009) Binding site detection and druggability index from first principles. *J. Med. Chem.* 52, 2363–2371

25 Guvench, O. and MacKerell, A.D., Jr (2009) Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comput. Biol.* 5, e1000435

26 Yang, C.Y. and Wang, S. (2010) Computational analysis of protein hotspots. *ACS Med. Chem. Lett.* 1, 125–129

27 Lexa, K.W. and Carlson, H.A. (2011) Full protein flexibility is essential for proper hot-spot mapping. *J. Am. Chem. Soc.* 133, 200–202

28 Bakan, A. *et al.* (2012) Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules. *J. Chem. Theory. Comput.* 8, 2435–2447

29 Antunes, D.A. *et al.* (2015) Understanding the challenges of protein flexibility in drug design. *Expert Opin. Drug Discov.* 10, 1301–1313

30 Wang, J.C. and Lin, J.H. (2013) Scoring functions for prediction of protein–ligand interactions. *Curr. Pharm. Des.* 19, 2174–2182

31 Lopes, P.E. *et al.* (2015) Current status of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* 1215, 47–71

32 Chen, J. and Brooks, C.L., 3rd (2007) Can molecular dynamics simulations provide high-resolution refinement of protein structure? *Proteins* 67, 922–930

33 Korb, O. *et al.* (2014) The cloud and other new computational methods to improve molecular modelling. *Expert Opin. Drug Discov.* 9, 1121–1131

34 Towns, J. *et al.* (2014) XSEDE: accelerating scientific discovery. *Comput. Sci. Eng.* 16, 62–74

35 Monticelli, L. and Tieleman, D.P. (2013) Force fields for classical molecular dynamics. *Methods Mol. Biol.* 924, 197–213

36 Vanommeslaeghe, K. and MacKerell, A.D., Jr (2014) CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta* 1850, 861–871

37 Tan, Y.S. *et al.* (2014) The use of chlorobenzene as a probe molecule in molecular dynamics simulations. *J. Chem. Inf. Model.* 54, 1821–1827

38 Raman, E.P. *et al.* (2013) Inclusion of multiple fragment types in the site identification by ligand competitive saturation (SILCS) approach. *J. Chem. Inf. Model.* 53, 3384–3398

39 Lakkaraju, S.K. *et al.* (2014) Sampling of organic solutes in aqueous and heterogeneous environments using oscillating excess chemical potentials in grand canonical-like Monte Carlo-molecular dynamics simulations. *J. Chem. Theory Comput.* 10, 2281–2290

40 Lakkaraju, S.K. *et al.* (2015) Mapping functional group free energy patterns at protein occluded sites: nuclear receptors and G-protein coupled receptors. *J. Chem. Inf. Model.* 55, 700–708

41 Kalenkiewicz, A. *et al.* (2015) Enrichment of druggable conformations from apo protein structures using cosolvent-accelerated molecular dynamics. *Biology (Basel)* 4, 344–366

42 Ciulli, A. (2013) Biophysical screening for the discovery of small-molecule ligands. *Methods Mol. Biol.* 1008, 357–388

43 Harner, M.J. *et al.* (2013) Fragment-based drug discovery using NMR spectroscopy. *J. Biomol. NMR* 56, 65–75

44 Foster, T.J. *et al.* (2012) Balancing target flexibility and target denaturation in computational fragment-based inhibitor discovery. *J. Comput. Chem.* 33, 1880–1891

45 Faller, C.E. *et al.* (2015) Site Identification by Ligand Competitive Saturation (SILCS) simulations for fragment-based drug design. *Methods Mol. Biol.* 1289, 75–87

46 Winn, M.D. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* 67, 235–242

47 Morris, G.M. *et al.* (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791

48 Yu, W. *et al.* (2014) Site-Identification by Ligand Competitive Saturation (SILCS) assisted pharmacophore modeling. *J. Comput. Aided Mol. Des.* 28, 491–507

49 Yu, W. *et al.* (2015) Pharmacophore modeling using site-identification by ligand competitive saturation (SILCS) with multiple probe molecules. *J. Chem. Inf. Model.* 55, 407–420

50 Raman, E.P. *et al.* (2011) Reproducing crystal binding modes of ligand functional groups using site-identification by ligand competitive saturation (SILCS) simulations. *J. Chem. Inf. Model.* 51, 877–896

51 Kozakov, D. *et al.* (2015) Ligand deconstruction: why some fragment binding positions are conserved and others are not. *Proc. Natl. Acad. Sci. U. S. A.* 112, E2585–E2594

52 Raman, E.P. *et al.* (2012) Site-specific fragment identification guided by single-step free energy perturbation calculations. *J. Chem. Theory. Comput.* 8, 3513–3525

53 Yang, C.Y. and Wang, S. (2011) Hydrophobic binding hot spots of Bcl-xL protein–protein interfaces by cosolvent molecular dynamics simulation. *ACS Med. Chem. Lett.* 2, 280–284

54 Ung, P.M. *et al.* (2016) Identifying binding hot spots on protein surfaces by mixed-solvent molecular dynamics: HIV-1 protease as a test case. *Biopolymers* 105, 21–34

55 Lexa, K.W. *et al.* (2014) Parameter choice matters: validating probe parameters for use in mixed-solvent simulations. *J. Chem. Inf. Model.* 54, 2190–2199

56 Tan, Y.S. *et al.* (2012) Using ligand-mapping simulations to design a ligand selectively targeting a cryptic surface pocket of polo-like kinase 1. *Angew. Chem. Int. Ed. Engl.* 51, 10078–10081

57 Alvarez-Garcia, D. and Barril, X. (2014) Molecular simulations with solvent competition quantify water displaceability and provide accurate interaction maps of protein binding sites. *J. Med. Chem.* 57, 8530–8539

58 Huang, D. *et al.* (2014) Structured water molecules in the binding site of bromodomains can be displaced by cosolvent. *ChemMedChem* 9, 573–579

59 Cavalier, M.C. *et al.* (2016) Small molecule inhibitors of Ca(2+)-S100 B reveal two protein conformations. *J. Med. Chem.* 59, 592–608

60 Samadani, R. *et al.* (2015) Small-molecule inhibitors of ERK-mediated immediate early gene expression and proliferation of melanoma cells expressing mutated BRaf. *Biochem. J.* 467, 425–438