Published in final edited form as:

Adv Exp Med Biol. 2010 ; 680: 481–488. doi:10.1007/978-1-4419-5913-3_54.

Docking to large allosteric binding sites on protein surfaces

Ursula D. Ramirez, Faina Myachina, Linda Stith, and Eileen K. Jaffe

Fox Chase Cancer Center, 333 Cottman Ave, Philadelphia, PA 19111

Small molecule docking to protein targets was developed as a drug discovery tool at a time when drug discovery was focused predominantly on enzyme active sites rather than allosteric sites. However, as early as 1963, Monod and coworkers astutely pointed out that protein function can be regulated allosterically by small molecules that need not be structurally related to a protein's target (e.g. substrate) because they bind somewhere other than the active site [1]. Some allosteric sites are similar to active sites, which are often located in deep grooves, somewhat buried in the protein structure, and contain a limited set of residues that provide a well defined binding site. Other allosteric sites are more like protein-protein interfaces (PPIs), which have less rigid binding requirements, frequently have species-specific variations in composition, and have overall greater structural flexibility.

Application of computational docking to allosteric sites or PPIs has been done successfully [2–5] despite the fact that most related software has been developed and benchmarked with active sites in mind [6]. Oftentimes targeting PPIs is undertaken to find small molecules that will interfere with oligomer assembly. Alternatively, one can target a specific cleft formed at a PPI with the notion of stabilizing a particular assembly. We have undertaken a series of docking studies that address allosteric sites that occur at PPIs where small molecule binding modulates an equilibrium of functionally distinct alternate quaternary structure assemblies [4,5]. Proteins with such assemblies have been called morpheesins and the equilibrium can be illustrated using a morphing dice schematic (Fig. 1). The distinguishing characteristic of proteins that function as morpheesins is that there exist alternate assemblies and each oligomeric assembly may have surface clefts that are assembly-specific. These clefts do not have the evolutionary requirement for conservation that is characteristic of active sites. The utility of docking to clefts in PPIs has been established [7,8]. Oligomer-stabilizing, small molecule binding to one oli-gomeric form of a morpheein equilibrium can be schematically represented in the dice analogy by a tetramer-specific wedge whose binding draws the equilibrium toward that oligomer (Fig. 1).

PBGS: the prototype morpheein drug discovery target

Porphobilinogen synthase (PBGS, EC 4.2.1.24), also known as 5-aminolevulinic acid dehydratase (or ALAD), catalyzes the first common step in the biosynthesis of the tetrapyrrole pigments (heme, chlorophyll, vitamin B₁₂). As such, PBGS is an essential enzyme for most organisms. The residues of the PBGS active site are highly conserved [9], and thus do not provide sufficient structural variation to yield species-selective inhibitors through computational docking. However, PBGS from some species are established to exist in a dynamic equilibrium of oligomeric forms including a high-activity octamer, a low-activity hexamer, the interconversion of which occurs at the level of a dimer whose conformation dictates the stoichiometry and architecture of the higher assembly state [4,5,10–14]. One dimer conformation is competent for assembly to the octamer and the other for assembly to the hexamer. The percentage of each component in the equilibrium of PBGS quaternary structure assemblies is dependent on protein sequence and responds to protein concentration, pH, and ligand binding at the active site or allosteric site [13–15]. The physiological relevance of the equilibrium of quaternary structure assemblies for human PBGS is established through the

relationship between this equilibrium and the disease ALAD porphyria [16]. The physiological relevance of the quaternary structure assemblies for plant PBGS is established by the existence of a naturally occurring allosteric activator of the plant protein [10]. Of importance to a discussion of docking is that a hexamer-specific cavity exists, which is not phylogenetically conserved. Therefore this cavity can be targeted for development of species-selective inhibitors as lead compounds for antimicrobials, herbicides, or pesticides. The structural basis for this binding site is described below.

The first crystal structures of PBGS revealed each subunit to have two domains, an $\alpha\beta$ -barrel and an extended N-terminal arm [17]. Crystal structures of the octameric assembly have been solved for PBGS from multiple species [18]. In each dimeric unit cell of octameric PBGS, the arm of one subunit wraps around the $\alpha\beta$ -barrel of the neighboring subunit making a hugging-dimer (Fig. 2) [17]. There is only one structure of a PBGS hexamer, which is of the naturally occurring (and porphyria-associated) human PBGS variant F12L [10]. The asymmetric unit of the hexameric PBGS crystal structure is a dimer with the N-terminal arms extended (“detached-dimer”) (Fig. 2). The conformation of the N-terminal arm determines which assembly can be formed (Fig. 2). The hexamer is less active than the octamer because it lacks an arm-to-barrel interaction which stabilizes the closed conformation of the active site lid [19].

The conformational differences between the subunits comprising the hexamers versus octamers define an oligomer-specific target region that can be utilized to find small molecules that preferentially bind only to the hexamer. We have targeted a hexamer-specific surface cavity to find small molecules that will trap the low-activity hexameric form of human and pea PBGS with some success. In the case of human PBGS, we docked to the crystal structure of the hexamer [4,10]; in the case of pea PBGS we docked to a homology model that was prepared using multiple template structures [5]. In that case, the human hexamer structure (PDB code IPV8) was used alongside the octameric *Pseudomonas aeruginosa* PBGS (*Pa*PBGS) structure (PDB code 1GZG [20]) to form a model of the *Pa*PBGS hexamer. The *Pa*PBGS model was used as the template to prepare homology models of pea and *Yersinia enterocolitica* PBGS (*Ye*PBGS) hexamers. Here we report the results of docking to the *Ye*PBGS homology model. We discuss the challenges and solutions we have found for these docking studies.

Computational docking studies for discovery of compounds that stabilize the inactive PBGS hexamer

The goal of these studies was to find small molecules that would preferentially bind to and stabilize the inactive PBGS hexamer. In the case of the human protein, such compounds would inhibit the protein and potentiate diseases related to low PBGS activity. These diseases are ALAD porphyria and lead poisoning. In the case of the prototype plant pea PBGS such compounds, if species selective, could form the starting point for development of an herbicide. In the case of *Ye*PBGS such compounds provide starting points for a new class of antibiotic agents. In all cases we used the same version of the docking software GLIDE with libraries of small molecules from Life Chemicals, Inc. that the company had purported to be drug-like with a molecular weights of approximately 350–500 Da. Against pea and human PBGS, we used only a 65,000 compound subset of the company’s library, whereas for *Ye*PBGS we used their entire “in stock” library (250,000 compounds). Here we report results for *Ye*PBGS and compare the hit compounds to those obtained from docking to the human and pea proteins.

A large docking site (a cubic region with 25Å dimensions into which the ligand must fit and a central cubic region with 14Å dimensions in which the center of the ligand must lie) was defined at the interface of three of the subunits that comprise the hexamer (Fig. 3a). Because the asymmetry in the PBGS dimer in the crystallographic asymmetric units impacts residues in the docking site, two different docking boxes were used as described previously [5]. We

docked to equivalent sites in both the human PBGS hexamer crystal structure and the pea PBGS and *Ye*PBGS hexamer homology models. Two dimensional structures supplied by Life Chemicals, Inc. were prepared for docking using LIGPREP (Schrödinger, Inc.) to limit molecules to those comprised of only C, H, N, O, S, P and halogens and then to generate 3 dimensional representations of all stereoisomers, tautomers, and ionic states within a narrow near-neutral pH range for each structure.

The same docking strategies were applied to all targets. Docking was done first by Standard Precision Mode (SP) of GLIDE version 3.5 docking software from Schrodinger, Inc [21]. The top 10% of compounds identified by SP Glidescore (a proprietary modified Chemscore) were then docked using Extra Precision (XP) Mode of GLIDE version 3.5. The top 10% of compounds identified by XP Glidescore (or about 1% of our starting library) were then further analyzed. These compounds were evaluated for solubility by calculation of Log S solubility estimate using QIKPROP (Schrödinger, Inc.), and those compounds with Log S less than -6 were eliminated from further consideration since compounds need to be soluble for testing in *in vitro* assays. Further consideration limited the remaining docked structures to those that were within proper distance to make van der Waals contacts or hydrogen bonds to at least one atom from each of the three subunits that comprise the docking site. This criterion limited compounds selected to those with the highest potential for stabilizing the hexamer. Finally the remaining compound poses were manually sorted to provide the broadest possible assortment of chemical diversity and binding locations within the docking region.

Because we do not know of naturally occurring ligands that bind to the targeted site, we have no *a priori* information about the chemical structure of the molecules for which we were searching. Thus, our approach was designed to maximize sampling of experimental space. Using the above approaches and a final manual by-eye inspection, approximately 100 compounds were selected for purchase and *in vitro* testing against each target. Based on supplier availability 76 compounds for pea PBGS, 77 for human PBGS, and 86 for *Ye*PBGS were obtained. Purchased compounds were tested using both native polyacrylamide gel electrophoresis to separate oligomeric forms of PBGS and activity assays to quantify inhibition of PBGS. As reported previously, out of twelve compounds seen to stabilize the human PBGS hexamer, chemical characterization yielded two compounds that significantly shifted the equilibrium of human PBGS, but not pea PBGS towards hexamer [4]. In the case of pea PBGS, out of 10 compounds that each provide some stabilization of the hexamer, only one was a potent inhibitor that shifted pea PBGS, but not human PBGS towards hexamer [5]. These results demonstrated that both homology models and crystal structures can be used to identify compounds by our methods. Here we present our results for *Ye*PBGS.

Hexamer stabilizing inhibitors of *Ye*PBGS

Eighty-six compounds were individually incubated at 37°C with *Ye*PBGS (1 mg/ml) and native gel electrophoresis was used to evaluate the percentage of hexamer and octamer present relative to protein that was incubated with solvent DMSO alone. This method determined whether the compounds caused an increased mole fraction of hexamer in the equilibrium of *Ye*PBGS oligomers. Incubations were done for 1 hour, 2 hours, 4 hours, 6 hours and 23 hours prior to loading gels. Five of the compounds (referred to here as ML-7C1, ML-7D1, ML-7H3, ML-7F3, and ML-7I5), showed some increase in the mole fraction of hexamer and a decrease in the mole fraction of octamer (Fig 3b). Compound ML-7C1 was selected from docking to one box and the other four compounds from the other docking box. The docked poses of the four compounds that docked to the same box are shown in Figure 3c along with the structures and chemical formulas of all five hexamer-stabilizing agents. Compound ML-7D1 showed maximal mole fraction hexamer by the end of the 2 hour incubation, compound ML-7C1 showed maximal mole fraction hexamer after 6 hours and the other three reached their maxima

by 23 hours (data are shown for 23 hour incubation only). Inhibition assay data were inconclusive for these compounds because the protein's activity was impaired by the extended incubation time with and without inhibitors (data not shown). Further characterization of these molecules requires that we overcome this experimental limitation.

The current work is significant in three ways. First, the hexamer-stabilizing compounds reported here offer a starting point for species-specific antibiotics. Although inhibition by these compounds could not be quantified, three of them (ML-7C1, ML-7D1, and ML-7H3) were shown to inhibit *Y. enterocolitica* growth in an *in vivo* disk zone inhibition assay (data not shown). Second, the results demonstrate the utility of docking to PPIs on homology models. Although PPIs have a lower phylogenetic conservation than active sites, PPIs also have significant structural flexibility. Residues in PPIs may sample many conformations of which the template crystal structure is only one. Thus, a homology model of a PPI is likely to represent a conformation that is populated in solution. Third, oligomer-stabilizing compounds may provide tools to assist in crystal structure determination, which in turn will allow a more critical evaluation of docking to homology models. The quaternary structure dynamic characteristic of morpheins has interfered with generation of crystals that diffract to high resolution. Co-crystal structures of PBGS with oligomer-stabilizing compounds are expected to allow a better understanding of docking to PPIs of homology models.

Acknowledgments

The authors acknowledge Susan Slechta for *in vivo* testing of inhibitors, the Fox Chase Cancer Center High Performance Computing Cluster, and grant support from the National Institutes of Health grants R01 ES003654 (EKJ), R21 AI063324 (EKJ), P30 CA006927 (FCCC), and T32 CA009035 (Institute for Cancer Research, a component of FCCC).

References

1. Monod J, et al. J Mol Biol 1963;6:306–29. [PubMed: 13936070]
2. Bond CJ, et al. Biochemistry 2000;39(50):15333–43. [PubMed: 11112519]
3. Gonzalez-Ruiz D, Gohlke H. Curr Med Chem 2006;13(22):2607–25. [PubMed: 17017914]
4. Lawrence SH, et al. J Biol Chem 2009;284:35807–17. [PubMed: 19812033]
5. Lawrence SH, et al. Chem Biol 2008;15(6):586–96. [PubMed: 18559269]
6. Alvarez, J.; Shoichet, B. Virtual Screening in Drug Discovery. 1. Boca Raton, FL; CRC Press: 2005. p. 470
7. Arkin MR, et al. Proc Natl Acad Sci U S A 2003;100(4):1603–8. [PubMed: 12582206]
8. Li Y, et al. Structure 2005;13(2):297–307. [PubMed: 15698573]
9. Jaffe EK. Chem Biol 2003;10(1):25–34. [PubMed: 12573695]
10. Breinig S, et al. Nat Struct Biol 2003;10(9):757–63. [PubMed: 12897770]
11. Kokona B, et al. Biochemistry 2008;47(40):10649–56. [PubMed: 18795796]
12. Selwood T, et al. Biochemistry 2008;47(10):3245–57. [PubMed: 18271513]
13. Tang L, et al. J Biol Chem 2006;281(10):6682–90. [PubMed: 16377642]
14. Tang L, et al. J Biol Chem 2005;280(16):15786–93. [PubMed: 15710608]
15. Jaffe EK. Trends Biochem Sci 2005;30(9):490–7. [PubMed: 16023348]
16. Jaffe EK, Stith L. Am J Hum Genet 2007;80(2):329–37. [PubMed: 17236137]
17. Erskine PT, et al. Nat Struct Biol 1997;4(12):1025–31. [PubMed: 9406553]
18. Berman HM, et al. Nucleic Acids Res 2000;28(1):235–42. [PubMed: 10592235]
19. Jaffe EK. Bioorg Chem 2004;32(5):316–25. [PubMed: 15381398]
20. Frere F, et al. J Mol Biol 2002;320(2):237–47. [PubMed: 12079382]
21. Halgren TA, et al. J Med Chem 2004;47:1750–59. [PubMed: 15027866]

Fig. 1. Cubic and tetrahedral dice illustrate the behavior of proteins that function as morpheins
 The quaternary structure dynamic characteristic of morpheins is illustrated using a dice analogy where one die can morph between cubic and tetrahedral shapes. The illustrated assemblies associate the die face with one spot with the die face with four spots such that the cubic die assembles into a tetramer and the tetrahedral die assembles into a pentamer. The analogy is that two different conformations of a protein subunit can each dictate assembly to a different oligomer. All dice in one assembly must be of the same shape. Thus, for example, the tetramer must come apart, and its component dice must change shape to a pyramid before participating in assembly into a pentamer. A diamond shape is used to illustrate a tetramer-specific stabilizing agent that can bind to the largest face of the tetramer. Binding of this shape would hinder tetramer dissociation, and thereby pull the equilibrium towards the tetramer.

Fig. 2. The PBGS quaternary structure equilibrium

The crystallographic asymmetric units of PBGS octamers and hexamers are hugging-dimer (middle-left) and detached-dimer (middle-right), respectively. These dimers assemble around a central axis to form the larger oligomers (90° rotation for the octamer (left) and 120° rotation for the hexamer (right)). The dimers are shown as ribbons with one subunit in a light shade and the other in a dark shade. Each dimer is shown in the context of its oligomer with the remaining subunits shown as spheres. The star illustrates the location of the hexamer-specific surface cavity.

Fig. 3. YePBGS hexamer-stabilizing compounds

a) Ribbon diagram of the three YePBGS hexamer subunits comprising the docking target; inner and outer docking boxes are indicated. The docked poses of compounds ML-7D1, ML-7F3, ML-7H3, and ML-7I5 are shown as sticks. b) YePBGS at 1mg/ml was not incubated (pre-incubation lane) or incubated for 23 hours at 37°C alone, with DMSO only or with one of the five YePBGS hexamer-stabilizing compounds (2 mM in DMSO). Samples were resolved on 12.5% polyacrylamide native PhastGels. c) Close-up of (a) with docked poses of compounds ML-7D1 in ball-and-stick, ML-7F3 in black wire, ML-7H3 light shade tube, and ML-7I5 in dark shaded tube. 2-Dimensional structural representations and molecular formulas of all five YePBGS hexamer-stabilizing compounds are shown.