Title
Homologous ligands accommodated by discrete conformations of a buried cavity

Journal
Proceedings of the National Academy of Sciences of the United States of America, 112(16)

ISSN
0027-8424

Authors
Merski, Matthew
Fischer, Marcus
Balius, Trent E
et al.

Publication Date
2015-04-21

DOI
10.1073/pnas.1500806112

Peer reviewed

Homologous ligands accommodated by discrete conformations of a buried cavity

Matthew Merski[1,2], Marcus Fischer[1], Trent E. Balius[1], Oliv Eidam[3], and Brian K. Shoichet[4]

Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158-2550

Edited by Brian W. Matthews, University of Oregon, Eugene, OR, and approved March 12, 2015 (received for review January 13, 2015)

Conformational change in protein–ligand complexes is widely modeled, but the protein accommodation expected on binding a congeneric series of ligands has received less attention. Given their use in medicinal chemistry, there are surprisingly few substantial series of congeneric ligand complexes in the Protein Data Bank (PDB). Here we determine the structures of eight alkyl benzenes, in single-methylene increases from benzene to *n*-hexylbenzene, bound to an enclosed cavity in T4 lysozyme. The volume of the apo cavity suffices to accommodate benzene but, even with toluene, larger cavity conformations become observable in the electron density, and over the series two other major conformations are observed. These involve discrete changes in main-chain conformation, expanding the site; few continuous changes in the site are observed. In most structures, two discrete protein conformations are observed simultaneously, and energetic considerations suggest that these conformations are low in energy relative to the ground state. An analysis of 121 lysozyme cavity structures in the PDB finds that these three conformations dominate the previously determined structures, largely modeled in a single conformation. An investigation of the few congeneric series in the PDB suggests that discrete changes are common adaptations to a series of growing ligands. The discrete, but relatively few, conformational states observed here, and their energetic accessibility, may have implications for anticipating protein conformational change in ligand design.

conformational change | protein–ligand complexes | congeneric series | homologous series | T4 lysozyme

The importance of conformational flexibility in protein–ligand interactions is widely acknowledged. Structural studies of model systems such as dihydrofolate reductase (1, 2), cyclophilin A (3), adenylate kinase (4), and others (5, 6) have suggested that conformational changes in the protein are coupled to progress along the catalytic reaction coordinate, and that local fluctuations can affect coupling between binding and global transitions (7). For signal transduction, the importance of such conformational changes has long been recognized (8), and has been emphasized by recent experimental (9) and computational studies (10). Accordingly, molecular dynamics simulations of protein–ligand complexes are now widely considered for ligand design (11–16).

Despite the attention lavished on protein conformational change overall, the incremental protein accommodations that might be expected over a series of ligand perturbations have received less consideration. Indeed, in the teeth of the "methyl, ethyl, propyl, butyl. . .futile" aphorism and the many medicinal chemistry programs that explore such incremental perturbations, surprisingly few crystal structures of congeneric ligands bound to a single protein are publicly available. Of the few there are, none resolve decisively how a protein might accommodate a congeneric series of ligands. If we define a congeneric series as one with at least six ligands related through an incremental change in functionality, then only 13 of these are known in the Protein Data Bank (PDB), and all but 2 (bold and underlined in *SI Appendix*, Table S4) of these undergo little conformational change upon ligand binding—a point to which we return. Conversely, ligand binding leads to substantial conformational changes in therapeutic targets such as aldose reductase (17), dihydrofolate reductase (2), and tRNA-guanine transglycosylase (18), but here the perturbations among the ligands have often not been systematic enough to disentangle changes in ligand size and polarity, making it harder to isolate the receptor conformational changes involved and their origins. In addition, similar ligands can adopt dissimilar binding modes in the same protein (19).

Ideally, one would like series of ligands where size and physical properties are increased incrementally without introducing other perturbations that could change binding determinants. Correspondingly, one would like a site where the growing ligand forces receptor accommodations. One system that recommends itself is the cavity site in T4 lysozyme created by the substitution Leu99→Ala (L99A cavity). Formed in the hydrophobic core of the protein, the resulting 150-Å^3 cavity is sequestered from solvent and is almost entirely apolar. Seminal studies by Matthews and colleagues demonstrated that this cavity can bind aryl hydrocarbons (20, 21), and since then the cavity and related mutants have become model systems for ligand recognition (22–29). Contributing to this status has been the commercial availability of thousands of likely ligands, many closely related to one another. This is something that is untrue of larger, more complicated binding sites, where fewer likely ligands are readily available, and fewer still in congeneric series.

Significance

Many medicinal chemistry programs change ligands incrementally to explore protein binding and to optimize binding affinity. How a protein accommodates such a growing ligand series has received remarkably little structural attention. Here we investigate eight congeneric ligands that grow by single-methylene additions, determining their protein-bound structures by X-ray crystallography, to investigate how a protein accommodates these changes. Rather than changing conformation smoothly to complement the ever-larger ligands, the protein site adopts a few discrete conformations as it expands. Inspection of the few other homologous series in the Protein Data Bank suggests that such discrete conformational adaptations to ligand binding are common, and may be an important consideration in ligand design.

Author contributions: M.M., M.F., T.E.B., and B.K.S. designed research; M.M., M.F., T.E.B., and O.E. performed research; T.E.B. contributed new reagents/analytic tools; M.M., M.F., T.E.B., O.E., and B.K.S. analyzed data; and M.M., M.F., T.E.B., O.E., and B.K.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The datasets reported in this paper have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4W51–4W59).

[1]M.M., M.F., and T.E.B. contributed equally to this work.

[2]Present address: Instituto de Biologia Molecular e Celular, Universidade do Porto, 4150-180 Porto, Portugal.

[3]Present address: pRED Informatics, Roche Pharma Research and Early Development, Roche Innovation Center Basel, 4070 Basel, Switzerland.

[4]To whom correspondence should be addressed. Email: bshoichet@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1500806112/-/DCSupplemental.

Here we determine the structures of eight alkyl benzenes in complex with the L99A cavity, including complexes with benzene, toluene, ethylbenzene, *n*-propylbenzene, *sec*-butylbenzene, *n*-butylbenzene, *n*-pentylbenzene, and *n*-hexylbenzene, as well as the apo cavity, at resolutions ranging from 1.39 to 1.80 Å. Because only benzene may be readily accommodated by the apo cavity but most other members of the series bound with greater affinity, this series seemed well-suited to exploring ligand-provoked conformational changes. We asked whether the cavity continuously adapted its conformation to these incremental enlargements in ligand size or whether instead the cavity jumped to discrete conformational states. Using multiconformational crystallographic refinement, we investigated whether the new conformations of the cavity were accessible to the ground state or were more distant in energy. Expecting continuous changes among conformations, we were surprised by the structures that emerged. Comparison with other series in the PDB suggests that these types of protein changes may be common, with implications for anticipating protein accommodation in ligand design.

Results

Structures for Congeneric Ligands Bound to L99A. An attraction of the L99A cavity was that its structure had already been determined in complex with a partly congeneric series of ligands, including benzene, ethylbenzene, and *n*-butylbenzene (21). To fill in the series, we determined structures with the missing toluene and *n*-propylbenzene and extended it with structures of *sec*-butylbenzene, *n*-pentylbenzene, and *n*-hexylbenzene. Structures were determined to 1.56, 1.64, 1.63, 1.80, and 1.39 Å, respectively (*SI Appendix*, Table S1). Unlike the earlier members of this series, which had been collected on a home X-ray source, these datasets were collected at a third-generation synchrotron and were refined as a multiconformer model using the program PHENIX (30, 31). For each complex, there was clear evidence for the ligands and surrounding cavity residues in unrefined $F_o - F_c$ electron density maps (Fig. 1 and *SI Appendix*, Fig. S1). On refinement of the *n*-propyl-, *sec*-butyl-, *n*-pentyl-, and *n*-hexylbenzene complexes, two major conformations of the protein were distinguishable in the region of the "flexible" F helix of the enzyme (residues 107–115), which gates the cavity site (Fig. 1 and *SI Appendix*, Fig. S1).

The movement of the F helix, which transitions from an α-helix toward a 3_{10} helix in these structures, contrasted with what had been observed in the earlier benzene, ethylbenzene, and *n*-butylbenzene structures, determined over 20 y ago. In the earlier structures, the F helix was not refined to adopt alternate states, which is reflected by a sharp rise in crystallographic B factors in those structures (*SI Appendix*, Fig. S2). This discrepancy motivated us to redetermine these structures afresh. We did so for the L99A protein in its apo state and in complex with benzene, ethylbenzene, and *n*-butylbenzene to resolutions of 1.45, 1.50, 1.79, and 1.68 Å, respectively, all at a third-generation synchrotron (*SI Appendix*, Table S1). In the four redetermined structures, the positions of the ligands were observed clearly in unrefined $F_o - F_c$ electron density, as were the positions of surrounding cavity residues (*SI Appendix*, Fig. S1). Here again, two major conformations of the protein were refined with convincing electron density (Fig. 1).

Protein Conformational Changes on Ligand Perturbation. Analysis of the nine structures, from the apo through to the complex with *n*-hexylbenzene, revealed three major conformations of the cavity, two of which typically existed in the same structure (Figs. 1 and 2). In the apo and in the benzene-bound structures a "closed" conformation of the cavity dominated, in which the ligand was fully enclosed by the cavity without obvious access to bulk solvent (Fig. 2). This closed cavity occupies ∼90% of the observed electron density; the remaining 10% was unmodeled (Figs. 1 and 2*A*). However, even in the toluene structure, a second, "intermediate" conformation of the cavity, which partially opens to bulk solvent (Fig. 2*B*), was observed at 20% occupancy (Fig. 2*A*). In this intermediate conformation, the hydrogen-bonding pattern changes from an α-helix to a shorter

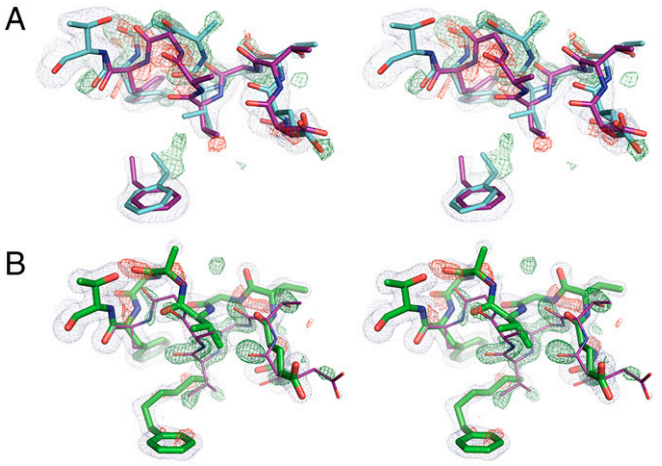

Fig. 1. Electron density maps reveal alternative F-helix conformations in the L99A cavity (stereoview). For illustration, only the major conformation of the F helix and of the ligands was refined ($2mF_o - DF_c$ maps as blue mesh, 1σ). The stick size corresponds to relative occupancies to which the three alternative F-helix conformations ultimately refined: closed (purple), intermediate (cyan), and open (green), which coexist in the individual complexes. When the major F-helix conformation was refined at 100% occupancy, difference electron density ($mF_o - DF_c$ maps, ±3σ) appears for missing alternative conformations (green mesh), whereas partial occupancy of the major F-helix conformation is indicated by negative density (red mesh). (*A*) In the ethylbenzene complex, the presence of the intermediate conformation of the F helix, and a second conformation of the ligand, is indicated by positive electron density (green mesh). Correspondingly, the closed conformation of the F helix, refined at 100% occupancy, is associated with negative electron density (red mesh); both conformations ultimately refined to around 50% (cf. Fig. 2). (*B*) In the *n*-hexylbenzene complex, the difference density maps support the presence of the closed conformation, which cannot accommodate the large ligand and is modeled as apo and the open conformation of the F helix; these ultimately refined to 30% and 70%, respectively.

length typical of a 3_{10} helix between the carbonyl oxygen of Thr109 and residue Ala112. Also, the hydrogen bonds from the carbonyl oxygens of Gly107, Gly110, and Gly113 to the amide nitrogens of residues Val111, Gly113, and Thr115, respectively, present in the apo structure, are lost (*SI Appendix*, Fig. S3*A*). The reduction in the number of backbone hydrogen bonds in the intermediate conformation suggests a higher energy state than the apo, closed conformation (*SI Appendix*, Fig. S3). This intermediate, expanded conformation became more dominant from the toluene to *n*-butylbenzene cavity complexes, ranging from 20 to 60% occupancy, although it continued to coexist with the closed conformation of the cavity in its ligand-bound state (Fig. 2). In the *n*-butylbenzene complex a third "open" conformation of the cavity appeared (30% occupancy), coexisting with the intermediate and closed conformations at 60 and 10% occupancy, respectively (Fig. 2). In this open conformation, the 3_{10}-like hydrogen bond between Glu108 and Val111 is maintained from the intermediate conformation, whereas the hydrogen bonds present in the closed conformation from the amides of Gly113 and Thr115 reappear, although decreased and increased in connectivity by one residue, respectively (*SI Appendix*, Fig. S3). The open conformation, which dominates the *n*-pentylbenzene and *n*-hexylbenzene structures, has two fewer intraprotein hydrogen bonds than the closed conformation, and is now open to bulk solvent (Fig. 2*B*), exposing substantially more hydrophobic surface area than either the closed or the intermediate states, consistent with a higher energy conformation (*SI Appendix*, Fig. S3).

Discrete Protein Conformational States. Quantitative analysis of the nine structures supports their clustering into three discrete protein conformations. We measured the rms deviations among the main-chain atoms of the nine structures for F-helix residues 107–115, the only region of the protein that underwent substantial

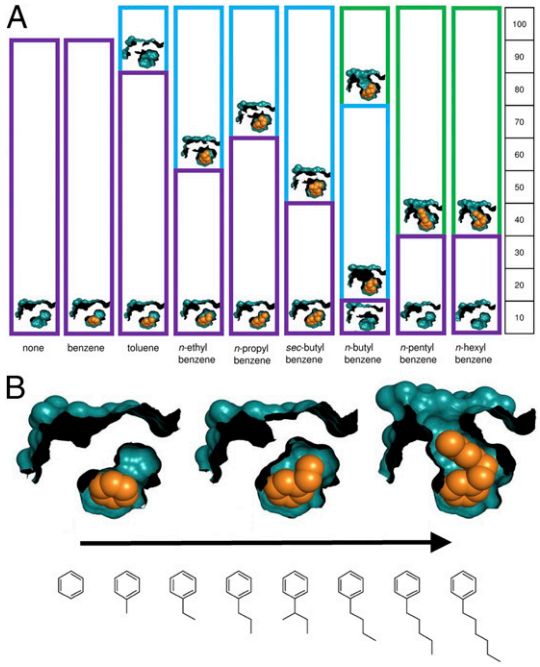

Fig. 2. Congeneric ligands are accommodated in L99A with conformational changes. (*A*) In the L99A cavity, the ligand poses were assigned to their respective protein conformations by matching the ligand occupancy with that of the F-helix conformation, which was typically unambiguous. (*B*) Molecular surface of the cavity, cut away to reveal the ligand (orange space-filling model), in examples of the closed (benzene complex), intermediate (ethylbenzene complex), and open (*n*-hexylbenzene complex) conformations. The full congeneric series is shown.

movement (Fig. 3 *A* and *B*). One conformation observed in each of the nine structures adopted a closed conformation, with rmsd values among the main-chain atoms within 1.4 Å of one another. For the apo structure and the benzene and toluene complexes this conformation dominated, whereas in the ethyl-, *n*-propyl-, and *sec*-butylbenzene structures, this closed conformation is about equiprevalent with the intermediate conformation (Fig. 2*A*). In the structures with the largest ligands, the closed conformation, which can still be observed as a minor component, likely represents the percentage of protein molecules that are unoccupied by ligand, and so are in fact in an apo state (*SI Appendix*, Table S6). Correspondingly, the intermediate conformation of the protein could also be clustered by main-chain rmsd values of typically 1.0 Å or less among all five structures in which it appeared. The three open conformations of the F helix also differed by no more than 1.0 Å (Fig. 3*B*). Conversely, between the closed, intermediate, and open states, structures differed from one another by rmsd values of at least 1.8 Å, typically more (Fig. 3 *A* and *B*). Both visually (Fig. 3*A*) and by measurement (Fig. 3*B*), the closed, intermediate, and open states may be distinguished, and represent jumps without substantial intermediates between them (see below for a partial exception). This clustering of states based on main-chain atoms could be replicated at the side-chain level. For instance, Val111, which in previous structures has been modeled to occupy multiple rotamers (24), also occupied three distinct states in the nine structures (Fig. 3 *C* and *D*). Whereas within each state slight variations in the position of Val111 were observed, these were uncorrelated with ligand size and seemed to differ little and without trend; the differences in the position of Val111 among the three states were unambiguous.

An exception to the occupancy of discrete states was the behavior of the main-chain atoms of Glu108 at the hinge region of the F helix. Whereas in the closed conformations these atoms cluster to occupy a single conformation (Fig. 3 *E* and *F*), a continuous and relatively smooth change is observed in the intermediate and open complexes, as the cavity expands to accommodate larger and larger ligands (Fig. 3 *E* and *F*). This hinge region is the one area of the structure where conformations varied smoothly as the ligands grew in size.

Energy of Ligand Binding and Conformational Strain. The appearance of multiple conformational states in a single structure suggests that the alternate conformations are relatively low-energy and accessible. To set boundaries on how high in energy these states might be, the ligand binding energies were compared with what might be expected if ligands were binding optimally, and not paying a cost for protein conformational change. As shown by Morton and Matthews in their seminal calorimetric study (21), as ligands grow from benzene to toluene, ethyl-, *n*-propyl-, and *n*-butylbenzene, affinity rises linearly, but at only half the pace of the water→octanol transfer free energies (Table 1) (the affinities of *n*-pentyl- and *n*-hexylbenzene were inaccessible owing to solubility limits). For instance, toluene binds to L99A with a K_d of 100 μM (ΔG_{bind} of −5.2 kcal/ mol), 0.3 kcal/mol better (lower) than does benzene (K_d of 175 μM, ΔG_{bind} of −5.5 kcal/mol) (21). Meanwhile, by water→octanol transfer energies, one might expect toluene to bind 0.8 kcal/mol better (20, 32), a difference between observed and expected affinities of 0.5 kcal/mol (Table 1). Similarly, *n*-butylbenzene binds 1.5 kcal/mol better than does benzene, whereas water→octanol transfer free energies would have it binding 2.9 kcal/mol better (20, 32). Several reasons for this difference in observed and expected binding energies have been mooted, including a protein reorganization energy (21). If we attributed all of the discrepancy between the observed $\Delta\Delta G_{bind}$ and that predicted by transfer free energies to the cost of changing conformations from the closed to the intermediate to the open states, then the maximum energy for the intermediate conformation, represented by the ethyl- and *n*-propylbenzene complexes, and for the open conformation, represented by the *n*-butylbenzene complex, cannot exceed 0.8 and 1.4 kcal/mol, respectively (Table 1). Thus, both the co-occupancy of the conformations in individual crystal structures and the retention of substantial ligand affinity for the larger protein cavity conformations suggest that the intermediate and open conformations represent relatively low-energy, accessible states.

Conformations in Other Lysozyme Cavities Recapitulate the Three States. To investigate the generality of the conformations observed among the nine structures determined here, we analyzed 121 lysozyme cavity structures in the PDB (including apo and ligand-complexed structures of L99G, L99A/M102Q, L99A/M102H, L99A/M102E, L99A/M102L, L99A/F153A, L99G/E108V, and L99A/E108V). With the exception of 11 of these, all of these structures were refined to represent a single structure, not allowing for multiple conformations to be occupied. Intriguingly, based on the rms deviations of the F helix, the same three states dominated these deposited structures as well (Fig. 3*G* and *SI Appendix*, Figs. S4–S7 and Tables S2 and S3). Eighty of these structures occupied the closed conformation, another 28 occupied the intermediate conformation, and 4 occupied the open conformation (*SI Appendix*, Figs. S4 and S5 and Table S2). Finding our three conformations in these independently determined structures attests to their accessibility and occupancy in response to ligands beyond our congeneric series.

Conformational Accommodations in Other Protein–Ligand Series. To investigate whether the discrete protein accommodations to the congeneric ligands are common or unusual, we searched the PDB for other ligand series (*Methods*). We sought series where each ligand was closely related to at least one other by a high topological similarity, and where there were six or more such ligands binding to a common protein; we insisted on six congeneric ligands to ensure that there were enough to observe trends. We applied two different criteria to select these series: (*i*) We looked at

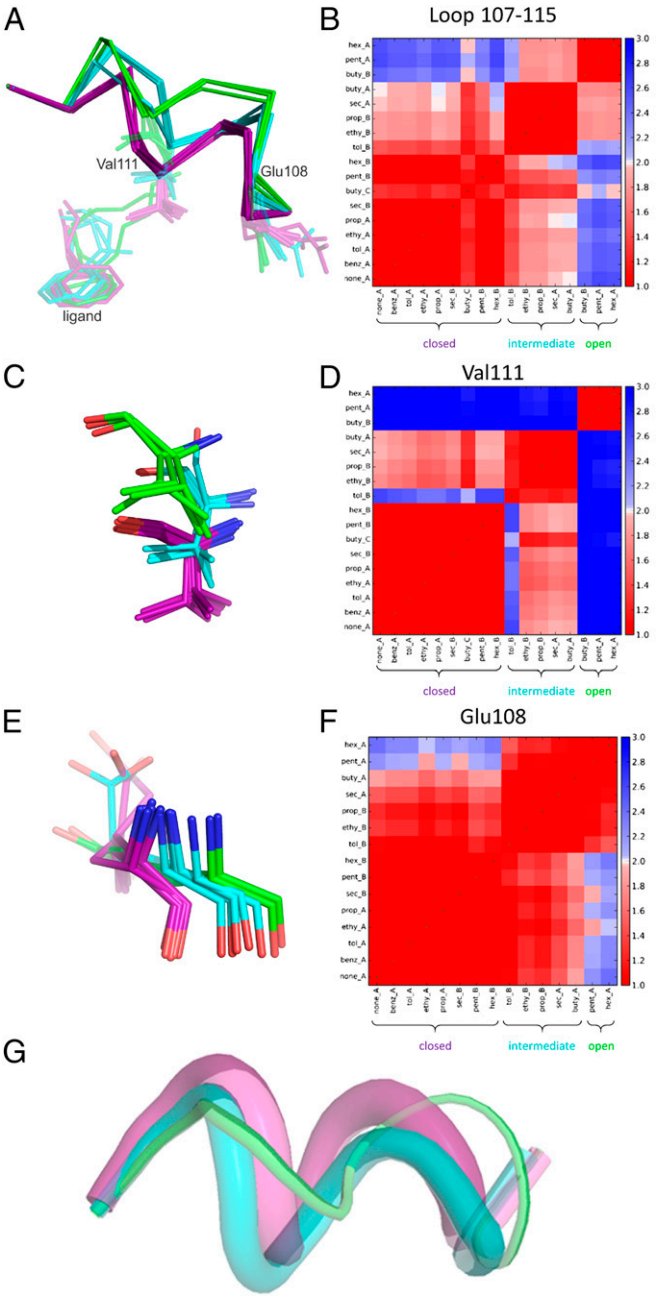

Fig. 3. Ligands are accommodated by mostly discrete conformational shifts of the L99A cavity. Structural superpositions (*A*, *C*, and *E*) and rmsd heat maps (*B*, *D*, and *F*) are shown for a ligand-responsive loop region (107–115; *A* and *B*) harboring Val111 (*C* and *D*) and Glu108 (*E* and *F*). Whereas most of the movement may be characterized as discrete (*C* and *D*), the Glu108 transitions (*E* and *F*) are at least partly smooth. Matrices *B*, *D*, and *F* are sorted by ligand size (closed, intermediate, and open; from small to large ligand size). Red indicates an rmsd value less than 2 Å, whereas blue indicates an rmsd value greater than 2 Å. (*G*) The three conformations of the cavity loop region (residues 107–115) are recapitulated among 121 crystal structures of cavity variants previously deposited in the PDB. The standard variations (Å) in the closed (purple), intermediate (cyan), and open (green) conformations in these structures are represented by tube width around the coordinate means of the clusters obtained by single-linkage hierarchical clustering.

the size of a linked list of ligands differing only by one heavy atom; and (*ii*) we weakened the linked list criteria and added a literature citation requirement. Because topological similarity is sensitive to ligand size (33, 34), we used the Tversky coefficient for fragments (70–250 Da) and the Tanimoto coefficient for larger molecules (70–500 Da), each with a minimum similarity coefficient of 0.6. To identify those series that were actually congeneric, we further insisted that they differed incrementally and linearly, interrogating each series closely (*SI Appendix*, Table S4). To our surprise, there were only 13 proteins that had congeneric series of six or more ligands in the PDB (*SI Appendix*, Table S4, underlined entries).

With the exceptions of the enzymes enoyl-ACP reductase (FabI) and Arg:Gly amidinotransferase, inspection of the protein structures reveals little monotonic response on the part of the protein to ligand binding. Often this reflected binding of the ligands on the surface of their proteins, allowing the ligand to grow into unfilled areas of the site and solvent (35–39). In the case of FabI, the enzyme responds to a series of six side-chain elongations of diphenyl ethers with a smooth shift of Ile207 and 0.5–0.9 Å movements of Tyr147 and Val201 (40) (*SI Appendix*, Fig. S8*E*). This is an example, then, of smooth side-chain change in response to a ligand perturbation, in contrast to the discrete changes we observed in the lysozyme cavity. For Arg:Gly amidinotransferase, there is evidence for both a smooth and discrete transition. For the discrete change the protein is in one state to bind glycine, γ-aminobutyric acid, and δ-aminovaleric acid; there is a conformational transition when it binds norvaline, alanine, or α-aminobutyric acid, with a corresponding gain of several hydrogen bonds (*SI Appendix*, Fig. S10).

When we relax our criteria, we find several other proteins that respond to smaller congeneric series of ligands with conformational accommodation. We analyzed four cases in detail: sialidase NanB (five congeneric ligands with structures), the estrogen receptor (two sets of four congeneric ligands), heat shock protein HSP90 (three congeneric ligands), and dihydroorotate dehydrogenase (three congeneric ligands) (*SI Appendix*, Table S5). In each, the protein responds to growing ligands with discrete conformational accommodations. In NanB (41), the Ile350–Asn353 loop opens from a closed state upon binding 2-[(3-bromobenzyl)amino]ethanesulfonic acid and 2-[(4-methoxybenzyl) amino]ethanesulfonic acid (PDB ID codes 4FPE and 4FPY). Intriguingly, difference electron density for the structures of PDB ID codes 4FPY, 4FPE, and 4FQ4 suggests that both closed and open loop states are present in the respective other structure (*SI Appendix*, Fig. S8*B*). The presence of the alternative conformations in any single structure is not modeled in the deposited data; instead, as in the early L99A complexes, the B factors of this region have been allowed to rise to values over 90 $Å^2$ (*SI Appendix*, Fig. S8*C*). Similarly, the estrogen receptor responds to the larger benzoxathiins and chromanes (PDB ID codes 1XPC, 1SJ0, and 1YIM) with a more open, tamoxifen-like conformation of helices h3, h11, and h12, forming an antagonist-like conformation. A more closed conformation is observed for a smaller ligand series that was designed to probe the impact of dynamic ligand binding on cellular signaling pathways (42). In HSP90, accommodations are more entangled, as here all of the ligands are fragments, sometimes binding in different configurations, and although certainly related they are not congeneric (apart from purine-based inhibitors such as PU3, PU8, and PU9). Nevertheless, it is germane to note that the divergent helical region from Leu103 to Lys116 exhibits three discrete conformations: a closed conformation for ligands such as adenine, a more open state for 9-ethyl-9H-purine-6-ylamine, and again more open for more decorated purines such as the congeneric series mentioned above. Finally, discrete states may be observed for dihydroorotate dehydrogenase around Pro131 (residues 128–139) upon binding a series of tetrahydropyrimidine carboxylic acids. Rather than responding to changes in the homologous series, it appears that a 40–50° rotation of the ligand's head group (e.g., dimethoxyphenyl in ligand W75 of PDB ID code 3W75) populates an alternative, open state, reflected by a change in the hydrogen-bond pattern of the protein backbone (*SI Appendix*, Fig. S11). In the difference electron density maps, we found evidence that those states not only exist within separate chains of the

Table 1. Ligand affinity and protein conformation

Ligand	K_d, μM*	$\Delta\Delta G_{bind}$, kcal/mol*,†,‡	$\Delta\Delta G_{wat,oct}$, kcal/mol*,†	Occupancy, %		
				C	I	O
Benzene	175	0.0	0.0	90	—	—
Toluene	102	−0.33	−0.82	80	20	—
Ethylbenzene	68	−0.57	−1.40	50	50	—
n-Propylbenzene	18	−1.36	−2.13	60	40	—
n-Butylbenzene	14	−1.50	−2.91	10	60	30

C, closed; I, intermediate; O, open loop conformation.
*From refs. 20 and 35.
†$\Delta\Delta G$ with reference to benzene.
‡The ΔG of benzene is −5.19 ± 0.16 kcal/mol.

homodimer, as modeled, but that both states also coexist within one of the monomers (*SI Appendix*, Fig. S8*F*).

Discussion

Protein conformational change upon ligand binding has received much attention (43, 44); what sets this study apart is its focus on protein accommodations to a congeneric series of ligands. A steady perturbation of molecular properties is widely practiced in ligand design but has surprisingly few examples in published structures. Three principal observations emerge. First, to a congeneric series of eight alkyl benzenes, the lysozyme L99A cavity responds by discrete conformational changes, transiting among three states. Within each state there were small variations, revealing no steady pattern, whereas the closed, intermediate, and open conformations of the cavity were readily distinguished and represented clear responses to ligand enlargement. Second, in multiconformational crystallographic refinement, two of these three states coexisted in most of the single complexes determined. This, and consideration of the ligand affinities, suggests that these alternative conformations represent not only discrete but also low-energy conformations. Third, a review of PDB structures suggests that discrete protein accommodations to congeneric ligands are not uncommon, transiting among relatively few apparently low-energy protein conformations, as far as we can determine given the few systematic cases known. These observations have implications for anticipating protein accommodations in ligand design.

Both structural analyses and binding energy support the idea that the lysozyme cavity responds with discrete, low-energy conformational changes as the alkyl benzenes grow in size. The conformations adopted by the F helix, as it responds to the ligands, fall into three discernible groups of main-chain conformations (Fig. 3). Within each of these groups are small variations, but these do not themselves track with an increase in ligand size. What does track with increasing ligand size is the discrete and progressive opening of the cavity in three clusters of conformations. The energetic accessibility of the more open conformations is supported by two lines of evidence. First, two or more conformations coexist with one another in all but one of the complexes determined here. Second, if one attributes all of the "missing" ligand binding energy, expected from increasing hydrophobicity, to protein reorganization (Table 1), then the cost of accessing the open conformation is not more than 1.4 kcal/mol; we consider caveats to this argument below.

If the closed, intermediate, and open conformations are accessible conformational states of the lysozyme cavity, then they should be seen in other cavity structures, responding to different perturbations. Consistent with this view, 121 previously published lysozyme cavity structures, representing cavity variants and their complexes with multiple ligands, could be readily clustered into the same three conformational substates represented by the nine structures determined here (*SI Appendix*, Fig. S4). These earlier structures typically represent single structure refinements that nevertheless recapitulate the same major states.

Similarly, if discrete protein accommodation to congeneric ligands is common, then it should be observable in other protein–ligand structures. Although we were surprised at just how few substantial congeneric series are represented structurally, the few there are often do undergo such discrete conformational changes. Thus, NanB, the estrogen receptor, dihydroorotate dehydrogenase, HSP90, and Arg:Gly amidinotransferase all respond to congeneric series of ligands with discrete conformational accommodations. As with the L99A cavity, these discrete conformations could often be observed in the same crystal structure when the electron density is inspected, even if, as was the case with NanB, they were originally modeled as having only a single, typically high B-factor conformation (21, 41). Naturally, we do not pretend that such discrete conformational accommodations are the only ways for a protein to respond to congeneric ligands; there were also several cases where the proteins responded to the ligands relatively smoothly, as was the case for FabI and Src kinase, and several congeneric series could be accommodated without movement of the protein at all. Still, the movement observed in the protein was often a substantial, discrete one, as though jumping from one low-energy state to another.

Certain caveats merit airing. Even in the L99A cavity, there are movements that can be represented as smooth rather than discrete. For the motions of the main-chain atoms of Glu108, on the periphery of the F helix, the distinction between the open and intermediate conformations collapses into a smooth accommodation to larger ligands (Fig. 3 *E* and *F*); the closed conformation could still be distinguished as a separate state. Neither do we pretend to have undertaken a comprehensive study of protein accommodations to small ligand changes, although we do hope to have captured the responses to most congeneric series. In many complexes, a small ligand perturbation will lead to a correspondingly small change in the protein until a tipping point is reached that pushes the protein into discrete conformational change, to occupy what appear to be a relatively small number of alternate, low-energy states. The inference that the intermediate and open conformational states are low-energy, based on differential ligand binding energies, depends on what one expects for the binding energies unencumbered by protein conformational strain. If optimal binding affinity should exceed the contribution of increased ligand hydrophobicity, then the inferred conformational strain would rise. However, even if one insists on an atomic ligand efficiency of 1 kcal/mol per added methylene (the ligand efficiency of benzene in the L99A cavity is 0.85), then *n*-butylbenzene should bind 4 kcal/mol better than benzene, rather than the 1.5 kcal $\Delta\Delta G$ actually observed. If one attributed that difference entirely to protein conformational strain, then the energy difference between the closed state and the more open states would still only be 2.5 kcal/mol.

Ligand optimization is the stage on which most effort is lavished in drug design and probe development. In such optimization, ligands are often perturbed incrementally. It is natural to assume that proteins will respond with correspondingly small accommodations, which may be modeled by techniques with small radii of convergence, such as structure relaxation and short molecular dynamics simulations. This study suggests that relatively large, discrete changes in protein conformation to small ligand perturbation may be common. Anticipating such changes may demand long molecular dynamics simulations (10, 45–48) or approaches that sample among precalculated states (49, 50). The low energy of these states would also support a combination of experimental observation, such as NMR or crystallographic refinement, with computational modeling of protein conformational changes in ligand optimization (31, 51). Irrespective of method, modeling discrete jumps among relatively few low-energy protein states may often be an important consideration in the structure-based design of new ligands.

Methods

Protein Crystallization and Structure Refinement. T4 lysozyme/L99A was cloned, purified, and crystallized as described (*SI Appendix*, *Methods*). Datasets on cryocooled crystals were collected at the Advanced Light Source beamline 8.3.1 and processed with XDS (52) and Phaser (53). To remove model bias from molecular replacement model 181L, F-helix residues 107–115

and the ligand were excluded from the starting model and added in the later rounds of model building, when occupancies were refined automatically using Phenix.refine (30) applying a 10% cutoff for modeling alternative conformations (cf. Fig. 2*A*). The presence of additional conformations in a structure was evidenced by features in the $F_o - F_c$ density maps ($\sigma = 3.0$) and decreases in R_{free} upon modeling additional conformations. Datasets were deposited in the PDB as 4W51 (no ligand), 4W52 (benzene), 4W53 (toluene), 4W54 (ethylbenzene), 4W55 (*n*-propylbenzene), 4W56 (*sec*-butylbenzene), 4W57 (*n*-butylbenzene), 4W58 (*n*-pentylbenzene), and 4W59 (*n*-hexylbenzene).

PDB Search for Other Homologous Series. We included protein systems that met the following criteria: (*i*) ≥6 linked ligands with Tanimoto index >0.6, and (*ii*) have a molecular formula difference of ≤1 heavy atom. As an alternative, we replaced criteria (*ii*) with ≥6 ligands associated with one paper. We performed this procedure for both ligands (70–500 Da) and fragments (70–250 Da) and calculated a (weighted) Tanimoto coefficient with a cutoff of 0.6. Corresponding lists of (*i*) 65 and (*ii*) 120 structures (40 overlap) were hand-curated for references, ligand similarity, and PDB structures. More details are in *SI Appendix, Methods*; Python and csh scripts are available upon request.

For the analysis of 121 PDB-deposited lysozyme cavity structures and for our 9 structures, we used UCSF Chimera (54) for the alignment and rmsd calculation of the F helix (residues 107–115). Hierarchical clustering analysis was performed and heat maps were generated using Python 2.7.

ACKNOWLEDGMENTS. We thank Matthew J. O'Meara and Sarah Barelier for reading this manuscript. For enlightening conversations on the lysozyme cavity, B.K.S. thanks Andy Morton (1964–1997). Supported by GM59957 (to B.K.S.) and National Research Service Award F32GM108161 (to T.E.B.).

1. Chen S, et al. (2013) Detection of dihydrofolate reductase conformational change by FRET using two fluorescent amino acids. *J Am Chem Soc* 135(35):12924–12927.
2. Schnell JR, Dyson HJ, Wright PE (2004) Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu Rev Biophys Biomol Struct* 33:119–140.
3. Fraser JS, et al. (2009) Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 462(7273):669–673.
4. Henzler-Wildman KA, et al. (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature* 450(7171):838–844.
5. Kokkinidis M, Glykos NM, Fadouloglou VE (2012) Protein flexibility and enzymatic catalysis. *Adv Protein Chem Struct Biol* 87:181–218.
6. Henzler-Wildman K, Kern D (2007) Dynamic personalities of proteins. *Nature* 450(7172):964–972.
7. Whitten ST, García-Moreno E B, Hilser VJ (2005) Local conformational fluctuations can modulate the coupling between proton binding and global structural transitions in proteins. *Proc Natl Acad Sci USA* 102(12):4282–4287.
8. Elgeti M, et al. (2013) Precision vs flexibility in GPCR signaling. *J Am Chem Soc* 135(33): 12305–12312.
9. Chung KY, et al. (2011) Conformational changes in the G protein Gs induced by the β2 adrenergic receptor. *Nature* 477(7366):611–615.
10. Dror RO, et al. (2013) Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* 503(7475):295–299.
11. Cole DJ, Tirado-Rives J, Jorgensen WL (2015) Molecular dynamics and Monte Carlo simulations for protein-ligand binding and inhibitor design. *Biochim Biophys Acta* 1850(5):966–971.
12. Hsu YH, et al. (2013) Fluoroketone inhibition of Ca(2+)-independent phospholipase A2 through binding pocket association defined by hydrogen/deuterium exchange and molecular dynamics. *J Am Chem Soc* 135(4):1330–1337.
13. Ranganathan A, Dror RO, Carlsson J (2014) Insights into the role of Asp79(2.50) in β2 adrenergic receptor activation from molecular dynamics simulations. *Biochemistry* 53(46):7283–7296.
14. Young T, Abel R, Kim B, Berne BJ, Friesner RA (2007) Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc Natl Acad Sci USA* 104(3):808–813.
15. Singh N, Frushicheva MP, Warshel A (2012) Validating the vitality strategy for fighting drug resistance. *Proteins* 80(4):1110–1122.
16. Singh N, Warshel A (2010) Absolute binding free energy calculations: On the accuracy of computational scoring of protein-ligand interactions. *Proteins* 78(7):1705–1723.
17. Sotriffer CA, Krämer O, Klebe G (2004) Probing flexibility and "induced-fit" phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations. *Proteins* 56(1):52–66.
18. Brenk R, Stubbs MT, Heine A, Reuter K, Klebe G (2003) Flexible adaptations in the structure of the tRNA-modifying enzyme tRNA-guanine transglycosylase and their implications for substrate selectivity, reaction mechanism and structure-based drug design. *ChemBioChem* 4(10):1066–1077.
19. Mattos C, Rasmussen B, Ding X, Petsko GA, Ringe D (1994) Analogous inhibitors of elastase do not always bind analogously. *Nat Struct Biol* 1(1):55–58.
20. Morton A, Baase WA, Matthews BW (1995) Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry* 34(27):8564–8575.
21. Morton A, Matthews BW (1995) Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: Linkage of dynamics and structural plasticity. *Biochemistry* 34(27):8576–8588.
22. Boyce SE, et al. (2009) Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J Mol Biol* 394(4):747–763.
23. Wei BQ, Baase WA, Weaver LH, Matthews BW, Shoichet BK (2002) A model binding site for testing scoring functions in molecular docking. *J Mol Biol* 322(2):339–355.
24. Mobley DL, Chodera JD, Dill KA (2007) The confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. *J Chem Theory Comput* 3(4):1231–1235.
25. Wang K, Chodera JD, Yang Y, Shirts MR (2013) Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J Comput Aided Mol Des* 27(12):989–1007.
26. Nucci NV, Fuglestad B, Athanasoula EA, Wand AJ (2014) Role of cavities and hydration in the pressure unfolding of T4 lysozyme. *Proc Natl Acad Sci USA* 111(38):13846–13851.
27. Ucisik MN, Zheng Z, Faver JC, Merz KM (2014) Bringing clarity to the prediction of protein-ligand binding free energies via "blurring." *J Chem Theory Comput* 10(3):1314–1325.
28. Lakkaraju SK, Raman EP, Yu W, MacKerell AD, Jr (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(6):2281–2290.
29. Bouvignies G, et al. (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* 477(7362):111–114.
30. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
31. Fischer M, Coleman RG, Fraser JS, Shoichet BK (2014) Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat Chem* 6(7):575–583.
32. Sangster J (1989) Octanol-water partition coefficients of simple organic compounds. *J Phys Chem Ref Data* 18(3):1111–1227.
33. Flower DR (1998) On the properties of bit string-based measures of chemical similarity. *J Chem Inf Comput Sci* 38(3):379–386.
34. Wang Y, Bajorath J (2009) Development of a compound class-directed similarity coefficient that accounts for molecular complexity effects in fingerprint searching. *J Chem Inf Model* 49(6):1369–1376.
35. Biela A, et al. (2012) Impact of ligand and protein desolvation on ligand binding to the S1 pocket of thrombin. *J Mol Biol* 418(5):350–366.
36. Eidam O, et al. (2012) Fragment-guided design of subnanomolar β-lactamase inhibitors active in vivo. *Proc Natl Acad Sci USA* 109(43):17448–17453.
37. Mecinović J, et al. (2011) Fluoroalkyl and alkyl chains have similar hydrophobicities in binding to the "hydrophobic wall" of carbonic anhydrase. *J Am Chem Soc* 133(35):14017–14026.
38. Rauh D, Klebe G, Stubbs MT (2004) Understanding protein-ligand interactions: The price of protein flexibility. *J Mol Biol* 335(5):1325–1341.
39. Krimmer SG, Betz M, Heine A, Klebe G (2014) Methyl, ethyl, propyl, butyl: Futile but not for water, as the correlation of structure and thermodynamic signature shows in a congeneric series of thermolysin inhibitors. *ChemMedChem* 9(4):833–846.
40. Chang A, et al. (2013) Rational optimization of drug-target residence time: Insights from inhibitor binding to the *Staphylococcus aureus* FabI enzyme-product complex. *Biochemistry* 52(24):4217–4228.
41. Brear P, Telford J, Taylor GL, Westwood NJ (2012) Synthesis and structural characterisation of selective non-carbohydrate-based inhibitors of bacterial sialidases. *ChemBioChem* 13(16):2374–2383.
42. Srinivasan S, et al. (2013) Ligand-binding dynamics rewire cellular signaling via estrogen receptor-α. *Nat Chem Biol* 9(5):326–332.
43. Cui Q, Karplus M (2008) Allostery and cooperativity revisited. *Protein Sci* 17(8):1295–1307.
44. Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? *Proteins* 57(3):433–443.
45. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE (2012) Biomolecular simulation: A computational microscope for molecular biology. *Annu Rev Biophys* 41:429–452.
46. Lane TJ, Bowman GR, Beauchamp K, Voelz VA, Pande VS (2011) Markov state model reveals folding and functional dynamics in ultra-long MD trajectories. *J Am Chem Soc* 133(45):18413–18419.
47. Lane TJ, Shukla D, Beauchamp KA, Pande VS (2013) To milliseconds and beyond: Challenges in the simulation of protein folding. *Curr Opin Struct Biol* 23(1):58–65.
48. Fenley AT, Muddana HS, Gilson MK (2012) Entropy-enthalpy transduction caused by conformational shifts can obscure the forces driving protein-ligand binding. *Proc Natl Acad Sci USA* 109(49):20006–20011.
49. Das A, et al. (2014) Exploring the conformational transitions of biomolecular systems using a simple two-state anisotropic network model. *PLOS Comput Biol* 10(4):e1003521.
50. Lei M, Zavodszky MI, Kuhn LA, Thorpe MF (2004) Sampling protein conformations and pathways. *J Comput Chem* 25(9):1133–1148.
51. Pfleger C, Gohlke H (2013) Efficient and robust analysis of biomacromolecular flexibility using ensembles of network topologies based on fuzzy noncovalent constraints. *Structure* 21(10):1725–1734.
52. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
53. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4): 658–674.
54. Pettersen EF, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.