# Detection of ligand binding hot spots on protein surfaces via fragment-based methods: application to DJ-1 and glucocerebrosidase

Melissa R. Landon · Raquel L. Lieberman · Quyen Q. Hoang ·
Shulin Ju · Jose M. M. Caaveiro · Susan D. Orwig · Dima Kozakov ·
Ryan Brenke · Gwo-Yu Chuang · Dmitry Beglov · Sandor Vajda ·
Gregory A. Petsko · Dagmar Ringe

**Abstract** The identification of hot spots, i.e., binding regions that contribute substantially to the free energy of ligand binding, is a critical step for structure-based drug design. Here we present the application of two fragment-based methods to the detection of hot spots for DJ-1 and glucocerebrosidase (GCase), targets for the development of therapeutics for Parkinson's and Gaucher's diseases, respectively. While the structures of these two proteins are known, binding information is lacking. In this study we employ the experimental multiple solvent crystal structures (MSCS) method and computational fragment mapping (FTMap) to identify regions suitable for the development of pharmacological chaperones for DJ-1 and GCase. Comparison of data derived via MSCS and FTMap also shows that FTMap, a computational method for the identification of fragment binding hot spots, is an accurate and robust alternative to the performance of expensive and difficult crystallographic experiments.

**Keywords** Fragment-based drug design · Structure-based drug design · Hot spot identification · DJ-1 · Glucocerebrosidase · Parkinson's disease · Gaucher's disease · Pharmacological chaperones

M. R. Landon · Q. Q. Hoang · S. Ju · J. M. M. Caaveiro ·
G. A. Petsko · D. Ringe (✉)
Department of Biochemistry, Rosenstiel Basic Medical Sciences
Center, Brandeis University, 415 South Street MS 029,
Waltham, MA 02454, USA
e-mail: ringe@brandeis.edu

M. R. Landon
e-mail: mlandon@brandeis.edu

R. L. Lieberman · S. D. Orwig
School of Chemistry and Biochemistry, Georgia Institute
of Technology, Atlanta, GA, USA

D. Kozakov · G.-Y. Chuang · D. Beglov · S. Vajda
Department of Biomedical Engineering, Boston University,
Boston, MA, USA

R. Brenke
Bioinformatics Graduate Program, Boston University, Boston,
MA, USA

*Present Address:*
J. M. M. Caaveiro
Physical Biochemistry Laboratory, Department of Medical
Genome Sciences, Graduate School of Frontier Sciences,
The University of Tokyo, Tokyo, Japan

## Introduction

The increasing application of fragment-based methods in drug discovery is motivated by the desire to develop smaller, ligand-efficient [1] therapeutics with more desirable pharmacological properties, such as bioavailability. To achieve commensurate levels of potency as compared to larger molecules, however, fragment-based strategies must seek to exploit hot spots, i.e., sub-sites of binding regions that contribute significantly to the free energy of ligand binding [2]. In the absence of either biophysical or SAR data, discernment of hot spots within a binding region is often not possible. Here we present the application of two fragment-based strategies, the multiple solvent crystal structures (MSCS) method and its computational alternative FTMap, that rely solely on three-dimensional structure for the identification of hot spots. In this study both techniques are applied to DJ-1 and glucocerebrosidase (GCase), therapeutic targets for Parkinson's and Gaucher's diseases, respectively.

Parkinson's disease (PD, [3]) is a neurodegenerative disorder that results in the impairment of motor skills and

cognitive abilities [4], eventually leading to death. Although most cases of PD are sporadic, multiple genes are implicated in the development of PD [5]. Among these is DJ-1, a recent target to emerge for the development of anti-PD agents [6]. DJ-1 consists of 189 amino acids and its function is currently unknown; however, several point mutations are associated with PD [7]. The best-characterized PD-associated mutation of DJ-1 is L166P, located on a C-terminal helix in the homodimer interface [8]. This mutant loses its ability to dimerize [9], resulting in structural instability and rapid degradation [10]. Hence, stabilization of DJ-1 may serve to prevent PD-associated degradation.

Gaucher disease (GD) is the most prevalent of the lysosomal storage disorders, i.e., those caused by deficient lysosomal glycolipid hydrolase activity [11]. GD affects one in 60,000 individuals in the general population and one in 800 in the Ashkenazi Jewish population [12]. Loss of activity of acid-$\beta$-glucosidase (GCase), the enzyme that hydrolyzes glucosylceramide in the lysosome, causes GD; current treatments include GCase enzyme replacement therapy (ERT, Cerezyme) or substrate reduction therapy (SRT, Zavesca). Missense mutations in GCase not localized to its active site exhibit defects in cellular trafficking [13], but do not abolish enzymatic activity of GCase [14]. Pharmacological chaperone (PC) therapy [15] aims to rescue partially active GCase variants, such as the N370S mutant, for lysosomal trafficking and activity. Although somewhat counterintuitive, enzyme inhibitors can increase steady-state cellular levels of active enzymes. To date, high affinity competitive inhibitors of GCase that bind the active site yet do not inhibit activity have been tested as PCs for GD [16–21]. One such pharmacological chaperone is isofagomine (IFG), a nanomolar inhibitor that binds the active site of GCase [22] and increases cellular trafficking of the N370S mutant to the lysosome [16, 22]. However, once in the lysosome, a stabilizing molecule that binds to an allosteric site on the mutant enzyme would be preferable in order to prevent competition with the substrate.

Here we aim to identify allosteric binding regions to serve as starting points for the development of pharmacological chaperones for DJ-1 and GCase. Two fragment-based methodologies validated previously for the detection of hot spot regions on protein surfaces were utilized in this study. Multiple solvent crystal structures (MSCS, [23–25]) were determined to identify consensus solvent binding regions on the surface of DJ-1 and GCase. In MSCS experiments a crystalline protein is exposed to a series of organic solvents. Previous studies utilizing MSCS have shown that consensus sites, i.e., regions of the proteins surface where multiple solvent molecules co-localize, are indicative of ligand binding regions [24]. In conjunction with the MSCS experiments, FTMap [26], a computational

fragment-based method for the detection of hot spots, was employed. Similar in spirit to MSCS, FTMap, an improved version of the computational solvent mapping (CSMap, [27–29]) algorithm, identifies regions of consensus binding of fragment-sized molecules [26]. Although similar programs such as GRID [30] and MCSS [31] are commonly utilized for the detection of binding regions, studies performed using both CSMap and FTMap have shown that in addition to identifying ligand binding regions with a lower false positive rate than other approaches, both methods accurately detect hot spots determined by biophysical methods such as calorimetry and NMR [26, 32]. Here we report multiple hot spots on the surfaces of DJ-1 and GCase suitable for ligand design. Furthermore, comparison of data resulting from the MSCS experiments to those derived using FTMap show that the computational method accurately reproduces experimental findings in addition to providing further elucidation of the binding regions detected via MSCS, providing a faster and cheaper alternative to crystallographic experiments.

## Methods

### Multiple solvent crystal structures of DJ-1 and GCase

*Purification and crystallization.* Recombinant human DJ-1 and GCase were expressed and purified to homogeneity according to published protocols [8, 22].

Crystallization of DJ-1 with solvents was accomplished using hanging drop vapor diffusion. The drops consisted of 2 μL concentrated protein (30 mg/mL) plus 2 μL mother liquor. The mother liquor was comprised of 30% polyethylene glycol (PEG) 400 for cryoprotection, 100 mM Tris–HCl pH 8.5, 200 mM sodium citrate and, individually, 20%(w/w) Acetonitrile, 20% Ethanol, 10% dimethyl formamide (DMF), 20% dimethyl sulfoxide (DMSO), 10% Phenol, or 5% trifluoroethane (TFE). Crystals with dimensions of approximately 200 μm × 200 μm appeared overnight at 20 °C.

Crystals of GCase were grown as described previously [22] using the hanging drop vapor diffusion technique. Crystals were crosslinked by placing a micro-bridge (Hampton Research) with 3 μl of a 25% glutaraldehyde solution (Sigma) in the well below a crystal-containing drop and incubating for 30–45 min. For methanol soaking, crosslinked crystals were then transferred briefly to a solution containing 50% methanol, 5 mM acetate pH 4.5 and 15% PEG 400, and flash cooled in liquid N2. For soaking with phenol, crystals were briefly soaked in a solution containing 23 mM phenol in 30% glycerol and 10 mM acetate buffer pH 4.5, and flash cooled in liquid N2.

*Data collection and structure determination.* Diffraction data for DJ-1 were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) beamline 11-3. GCase diffraction data were collected at the GM/CA-CAT beamline at the Advanced Photon Source (Darien, IL). Data sets for both proteins were processed with the software program HKL2000 [33]. Solvent-bound DJ-1 structures were solved by molecular replacement with the program MolRep [34] using apo DJ-1 ([8], PDB entry 1SOA) as a search model, and subsequent refinement was performed using Refmac [35], both of the CCP4 package [36]. GCase structures were solved by rigid body refinement in Refmac5 using a model derived from PDB entry 2NT0 [22] after removing all non-protein molecules. Model building for all DJ-1 and Gcase structures were carried out in Coot [37]. Ligand parameters for DJ-1 were obtained from the Dundee PRODRG2 server (http://davapc1.bioch.dundee.ac.uk/prodrg). Water molecules were added to the model of GCase after several rounds of fitting atomic models into electron density maps using Coot and including phosphates and carbohydrates. For the GCase structures each non-protein molecule was subsequently inspected and noted for additional $F_o$–$F_c$ different electron density. Methanol or phenol molecules were then modeled, as appropriate. Data processing and refinement statistics for DJ-1 and GCase are recorded in Tables 1 and 2, respectively.

## FTMap

The FTMAP algorithm samples billions of fragment positions on a dense translational and rotational grid. Positions are scored using an energy function that includes attractive and repulsive van der Waals terms, electrostatic interaction energy based on Poisson-Boltzmann calculations, a cavity term to represent the effect of nonpolar enclosures, and a structure-based pair-wise interaction potential. In spite of its relative complexity, the energy expression is written as sum of correlation functions with components defined on grids. This enables the use of the extremely efficient fast Fourier transform (FFT) correlation method for function evaluation [38]. The FTMAP algorithm consists of five steps as follows.

1. *Rigid-body fragment docking.* Sixteen fragments (Fig. 1) are used to probe the binding surface of a protein of interest. This fragment set includes solvent types commonly employed in MSCS experiments as well as common chemical groups; previous analyses have shown that this set is sufficient to delineate hot spots within binding pockets [32]. Only atomic coordinates of the protein of interest are utilized for mapping simulations, i.e., no information on the binding site is required. Prior to FTMap simulations, all bound ligands and water molecules are removed from crystal structures. For each fragment a special purpose rigid-body docking algorithm using the FFT correlation approach samples billions of conformations. The 2,000 best poses for each fragment are retained for further processing.

2. *Minimization and re-scoring.* The free energy of each of the 2,000 complexes generated in Step 1 is minimized using the CHARMM potential with the Analytic Continuum Electrostatic (ACE, [39]) model representing the electrostatics and solvation terms as implemented in version 27 of CHARMM [40] using the parameter set from version 19 of the program. The ACE model includes a surface area dependent term to account for the solute–solvent van der Waals interactions. Minimizations are performed using an adopted basis Newton–Raphson method. During the minimization the protein atoms are held fixed while the atoms of fragments are free to move.

3. *Clustering and ranking.* Using a simple greedy algorithm, minimized fragment conformations generated from Step 2 are clustered to create consensus sites. First, for each fragment type, the lowest energy conformer is selected. Fragment conformers within 3 Å RMSD of this

**Table 1** Data refinement and statistics for DJ-1 MSCS experiments

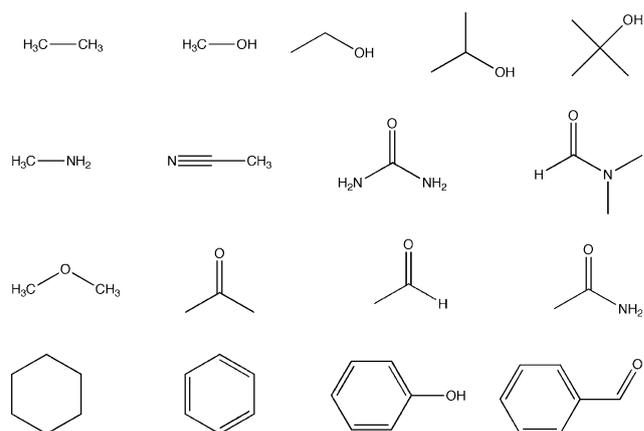| *Data collection* | | | | | |
|---|---|---|---|---|---|
| Solvent | ACN | Ethanol | DMF | DMSO | TFE |
| Space group | $P3_121$ | $P3_121$ | $P3_121$ | $P3_121$ | $P3_121$ |
| Wavelength (Å) | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 |
| Resolution (Å) | 1.53 | 1.53 | 1.53 | 1.53 | 1.53 |
| Completeness (%) | 99.77 | 99.82 | 99.93 | 99.79 | 99.92 |
| *Refinement* | | | | | |
| Unique reflection | 34,492 | 34,940 | 34,897 | 34,651 | 35,456 |
| R-factor ($R_{\text{free}}$) (%)[a] | 13.3 (15.6) | 11.8 (14.2) | 13.0 (15.7) | 15.7 (13.1) | 24.6 (25.7) |
| Average B-factor (Å$^2$) | 17 | 16.8 | 13.7 | 15.6 | 10.5 |

[a] R-factor $= \Sigma\|F_o|-|F_c\|/\Sigma|F_o|$, where $F_o$ and $F_c$ are the observed and calculated structure-factor amplitudes. $R_{\text{free}}$ is monitored with 5% reflections excluded from refinement

**Table 2** Data refinement and statistics for GCase MSCS experiments

| Bound solvent | Methanol | Phenol |
|---|---|---|
| *Data collection* | | |
| Space group | P2(1) | P2(1) |
| Cell dimensions | 108.4, 91.5, 152.5 | 113.2, 91.6, 153.2 |
| *a*, *b*, *c* (Å) | | |
| abg (°) | 90. 110.7, 90 | 90, 111.6, 90 |
| Resolution (Å)[a] | 20–2.40 | 15–2.30 |
| | (2.49–2.40) | (2.38–2.30) |
| $R_{sym}$ | 90(41.6) | 8.6(52.2) |
| I/s*I* | 16.1(2.6) | 24(2.8) |
| Completeness (%) | 93.2(79.8) | 89.3(95.7) |
| Redundancy | 3.0 | 3.5 |
| *Refinement* | | |
| Resolution (Å) | 20–2.40 | 15–2.30 |
| No. reflections | 78,243 | 109,635 |
| $R_{work}/R_{free}$[a] | 22.8/30.6 | 20.9/27.4 |
| No. atoms | | |
| Protein residues | 1,988 | 1,978 |
| *N*-acetyl-glucosamine (NAG) | 9 | 4 |
| Sulfate anion ($SO_4^{2-}$) | 14 | 9 |
| Glycerol (GOL) | | 4 |
| Water | 432 | 662 |
| MeOH | 5 | |
| Phenol | | 5 |
| *B*-factors | | |
| Protein | 45.5 | 39.6 |
| NAG | 66.9 | 53.5 |
| $SO_4^{3-}$ | 82.1 | 74.5 |
| GOL | | 52.6 |
| Water | 39.6 | 43.3 |
| MeOH | 52.4 | |
| Phenol | | 77.5 |
| R.m.s deviations | | |
| Bond lengths (Å) | 0.016 | 0.021 |
| Bond angles (°) | 1.865 | 2.016 |

[a] Highest resolution shell is shown in parenthesis. 5% of reflections were selected for $R_{free}$



**Fig. 1** Fragment set used in FTMap simulations

conformer are then joined to create the first cluster. An earlier analysis of binding distributions of fragments for several proteins revealed that fragments bind to the same subsite of a binding pocket with an average RMSD of 2 Å [41]. Here subsites are defined as regions that bind either an amino acid side chain (i.e., of a peptide substrate) or a comparably sized functional group of a ligand. Adjacent subsites are typically separated from each other by at least a 5 Å center-to-center distance; hence, in light of these findings, the value of 3 Å for the clustering radius is a conservative choice. After the first cluster is formed, the next lowest energy conformer is selected to start the second cluster. This step is repeated until all 2,000 poses are assigned to clusters. Clusters with less than six members are excluded from further consideration, thereby avoiding narrow energy minima with low entropy [42]. This cutoff has been determined empirically to result in a low false positive rate. Retained clusters are ranked on the basis of their Boltzman averaged energies.

4. *Determination of consensus sites.* Similar to MSCS experiments, FTMAP utilizes a consensus clustering approach to detect hot spots, i.e., regions of the protein surface where clusters of different fragment types overlap. The 96 clusters generated in the previous step (6 clusters for each of 16 fragment types) are used to form consensus sites. Using the distance between the centers of mass of the cluster centers as the distance measure, each fragment cluster is assigned to a consensus site. A clustering radius value of 4 Å is used since subsites are separated by greater than 5 Å (see Step 3). As in Step 3, FTMAP again employs a simple greedy algorithm to find the cluster with the maximum number of neighbors (defined as cluster centers within 4 Å from each other), which forms the first consensus site. Members of this site are then removed from consideration, and the procedure is repeated until all clusters are exhausted. Conformers are then redistributed among the consensus sites such that each conformer is closest to the center of its own consensus site, whereupon consensus sites are ranked based on cluster population size.

5. *Characterization of the binding site.* FTMap first selects the largest consensus site (CS1), which is generally identifies the most important subsite (or hot spot) for ligand binding. CS1 forms the kernel of the ligand-binding region. The binding site is then expanded by adding any consensus site (irrespective of its size) within 7 Å of any consensus site already in the binding site. This procedure continues until no further expansion is possible. The resulting set of consensus sites is used to describe the binding site.

## Results and discussion

Detection of hot spots for DJ-1

Experimental solvent mapping of DJ-1 revealed a single hot spot bound by all solvents on each monomer (Fig. 2a, circled). This hot spot emerges in the region containing Cys 106 and is shown in close-up in Fig. 2b. Published studies have shown that oxidation of Cys 106 results in the localization of DJ-1 to the mitochondria, where it is has a demonstrated neuroprotective effect [43]. Independent of the MSCS experiments, FTMap simulations (see "Methods") were conducted on a dimeric crystal structure of apo DJ-1 (PDB entry 1SOA, [8]). Several consensus regions for fragment binding are detected using FTMap (magenta, Fig. 3; Table 1). In excellent agreement with the MSCS data, the most highly populated consensus site resulting from FTMap coincides with the region surrounding Cys 106 (Fig. 3b). In addition to reproducing the experimental solvent mapping data, multiple consensus sites unique to the FTMap simulations are formed near the dimer interface (magenta sticks, Fig. 3a, b). These hot spots may provide a novel starting point for the development of stabilizing agents of DJ-1. In Table 3 the ranks of the consensus sites are given with their size, i.e., the number of fragment clusters comprising each site. Interacting residues are determined as those residues located within 5 Å of a fragment cluster and were calculated using Pymol version 0.99 ([44], http://pymol.sourceforge.net).

Detection of novel hot spots for GCase via MSCS and FTMap

Structures of cross-linked GCase soaked with methanol (Structure A) and phenol (Structure B) are shown here as
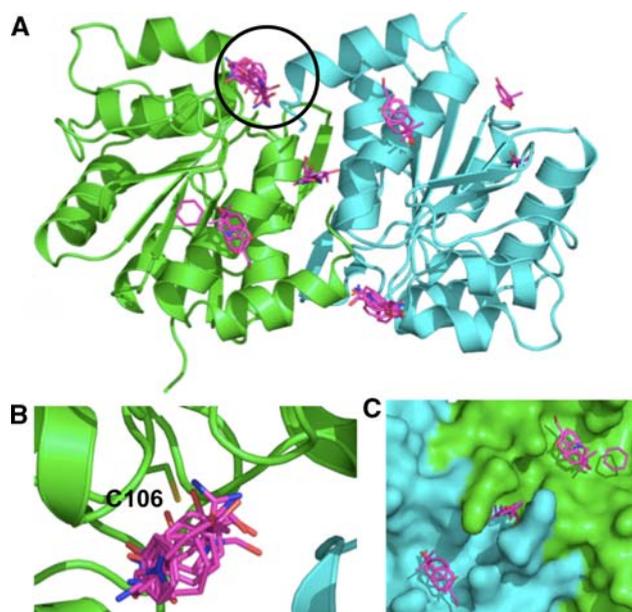


**Fig. 3** *Hot spots* detected for DJ-1 via FTMap simulations. **a** Fragment clusters (*magenta sticks*) populating the top six consensus sites resulting from the simulations are superimposed on dimeric DJ-1. The largest consensus site (*circled*) is located in the region also detected via MSCS. **b** Close-up view of *circled region* in (**a**). C106 is labeled. **c** Close-up view of additional hot spots formed close to the dimer interface. Based on these data, we hypothesize that a pharmacological chaperone could be targeted to this region

Figs. 4 and 5, respectively. Four molecules are present in the asymmetric units of both structures, where the individual monomers are nearly identical to previously solved structures ([22], not shown). Fourteen sulfate, and 12 methanol molecules were modeled in structure A, whereas nine sulfates and 18 phenol molecules were modeled in



**Fig. 2** Results of the DJ-1 MSCS experiments. **a** Solvents (*magenta sticks*, *circled*) bind to both monomers (*green* and *blue cartoon*) of DJ-1 in the region of cysteine 106. **b** Close-up view of the region surrounding the bound solvents. Cys106 is shown in *stick* to illustrate its proximity to the bound solvents

**Table 3** Summary of FTMap-derived hot spots for DJ-1

| CS ranks | CS size | Residues |
|---|---|---|
| 1, 2(B) | 23, 18(B) | E15, E18, R48, G74, N76, C106, A107, H126 |
| 3, 4(B), 5 | 13, 8(B), 5 | R145, G159, F162, E163, L187, K188, R145(B), R158(B), G159(B), F162(B), E163(B), L163(B), L187(B), K188(B) |

All results listed are for a single chain unless otherwise denoted
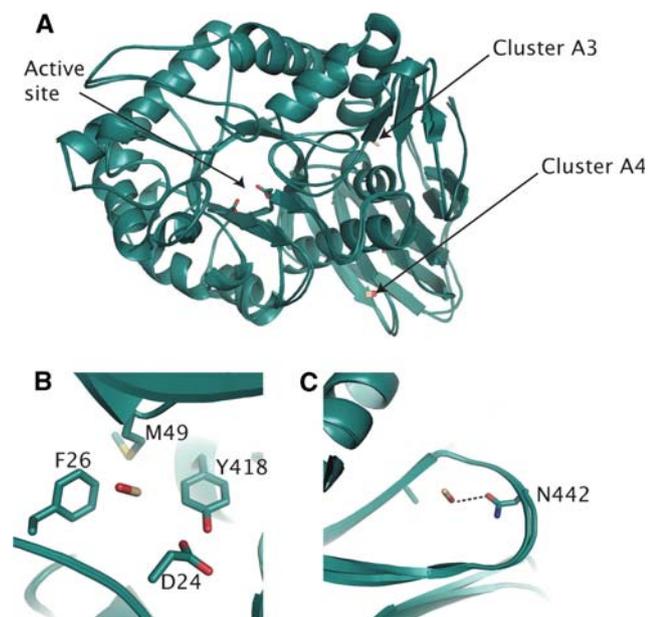
*CS* consensus site, *B* chain B



**Fig. 4** Bound methanols in GCase (Structure A). **a** Overall view of methanol clusters in GCase relative to active site. **b** Cluster A3, which is equivalent to cluster B in Fig. 6. **c** Cluster A4. Stabilizing interactions (<3.7 Å) between sulfate and side chains are shown with *dashed lines*

structure B. Four glycerol molecules, used as a cryoprotectant, were bound in the active site of structure A, a result we obtained previously and will not be discussed further in this analysis [22]. The modeled sulfate, phenol and methanol molecules are located throughout the surface of GCase (Figs. 4, 5, 6), and represent sites previously modeled with water and sulfates only. Clusters described in detail below are limited to those regions where a heteroatom of interest was modeled in at least two of the four monomers in the asymmetric unit of GCase. Inspection of the regions bound by the solvents yields several points are noteworthy. First, two tight-binding anionic sites are found in the structure A (Fig. 6). These clusters have been described previously [12, 45, 46], occupied by sulfate or phosphate anions in numerous other GCase crystal structures (not shown). Cluster A1 (Fig. 6b) resides on the same face of the TIM barrel domain as the active site, and sits at the end of the α-helix harboring N370, the site of the most common
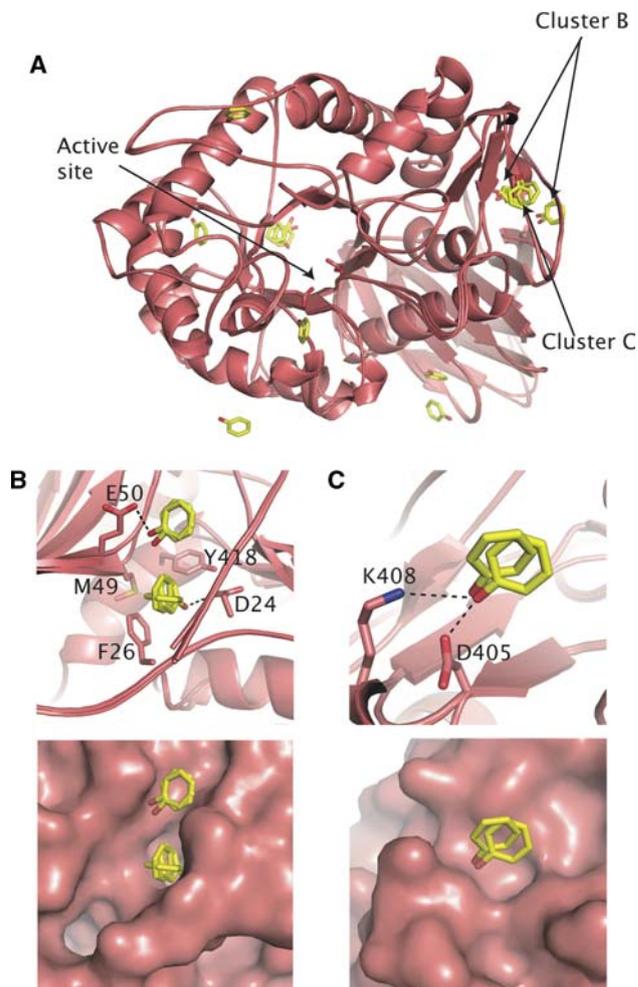


**Fig. 5** Bound phenols in GCase (Structure B). **a** Overall view of modeled phenols after superposition of all monomers in asymmetric unit. **b** Cluster B at interface of three secondary structural elements. *Top*, *ball*-and-*stick* view. *Bottom*, surface representation. **c** Cluster C at antiparallel β-strand domain. *Top*, *ball*-and-*stick* view. *Bottom*, surface representation. Stabilizing interactions (<3.7 Å) between sulfate and side chains are shown with *dashed lines*

mutation among Ashkenazi Jews [20]. The sulfate anion forms electrostatic/hydrogen bonding interactions with Arg 353 and Ser 12. Cluster A2 (Fig. 6c) is located at the opposite end of the TIM barrel from the active site, and is stabilized by interactions with Arg 277, Trp 228 and His
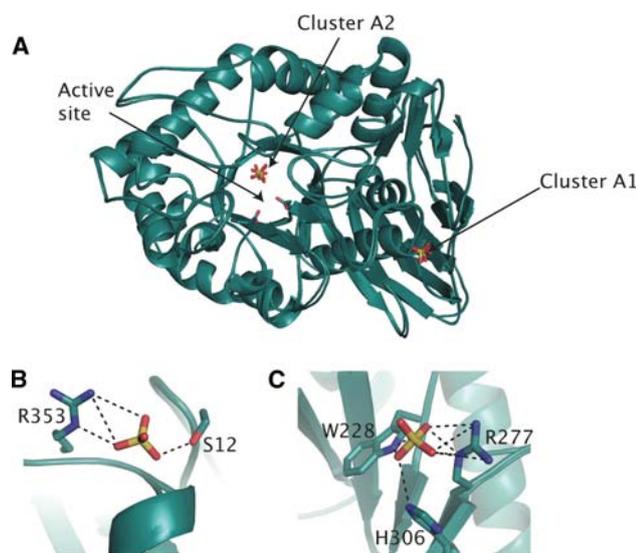
**Fig. 6** Bound sulfates in methanol-bound GCase (Structure A). **a** Overall view of the location of the clusters relative to the active site after superposition of all four monomers in the asymmetric unit. **b** Cluster A1 **c** Cluster A2. Stabilizing interactions (<3.7 Å) between sulfate and side chains are shown with *dashed lines*

306. The analogous site in structure B (Fig. 5a) is modeled as a mixture of sulfate and phenol. The observation that cluster A1 is best modeled with sulfate anion in both structures, even after cross-linking and soaking with organic solvents, provides experimental evidence for cluster A1 as a phospholipid binding site. This hypothesis has been suggested previously [22], and is consistent with the notion that a destabilized helix harboring the N370S mutation would impair phospholipid binding [47].

One cluster of interest is located at the interface of all three secondary structure elements in GCase, a groove shaped by amino acid side chains derived from the TIM barrel, β-barrel and anti-parallel β-sheet. In this site we have found phenol (Fig. 5b, cluster B), methanol (Fig. 4b, cluster A3), and previously, glycerol, which was introduced as a cryoprotectant (see for example PDB entry 2NT0). Another phenol appears bound close by in two of the four crystallographic monomers (Fig. 5b). A second site (Fig. 5c, cluster C) found on the anti-parallel β-sheet, is separated from the previous site by the N-terminus (residues 1–10). The functional significance of these sites is not yet clear, but it is possible to envision a scenario in which a signal triggers a conformational change in the N-terminus to expose this site for binding to a protein partner. Two partners are currently known: LIMP-II [48], which GCase is proposed to traffic to the lysosome, and saposin C, an activator protein [49–51]. However, the precise binding site for either of these proteins is not currently known. Finally, we have identified a cluster of methanol molecules in structure A (Fig. 4c, cluster A4) at

the edge of the β-barrel that abuts the TIM barrel, particularly at the N370-containing helix (see above). N370 has been shown to be critical in stabilizing an active site loop in a conformation of GCase that substrate-ready [22]. It may be possible to develop a small molecule stabilizer for N370S GCase at this interface, which would both improve cellular trafficking as well as improve lysosomal catalysis. This region is also a candidate for a binding site for saposin C, as N370S mutant GCase exhibits poor binding to this activator protein [47]. Further functional studies will be required to determine the binding surface on GCase for both Sap C and LIMP-II, but these experimental methods provide clues to sites of interest.

Computational mapping simulations using FTMap were performed on three GCase monomers (Fig. 7; Table 4). These monomers were chosen to ensure that all published conformations of GCase were sampled for this analysis. For this study two monomers of PDB entry 2NSX [22] were utilized: chain B, to which IFG is bound, and chain A, to which only glycerol is bound. When IFG is bound, a conformational change in loop 1 occurs, resulting in a shallower surface topology as compared to the apo structure [22]. The third structure utilized was the A chain of PDB entry 2NTl, a neutral pH apo structure of GCase where no bound ligands are present. Two regions containing well-populated consensus sites are found in all three structures [Fig. 7, circled in (A)]. The largest concentration of fragment clusters contains the catalytic center and proposed hydrophobic subsites for binding of the substrate, glucosylceramide (large circle in Figs. 7a, 8). A comparison of the consensus sites for glycerol-bound GCase (2NSX chain A) and apo GCase (2NT1 chain A) shows high overlap at the site of catalysis (Fig. 8a). In addition to this region, a second cluster is observed only in glycerol-bound GCase (Fig. 8a). Some of these regions appear to clash with the protein side chains in GCase, but are clarified in a comparison with the IFG-bound structure (Fig. 8a). Overlay of all hot spots formed in the catalytic region point to a binding region that encompasses both the site of catalysis and the two emerging troughs; these data lend further credibility to the hypothesis that the induced fit conformation adopted by bound IFG is the substrate-ready form of the enzyme. Further, the hot spots that clash with GCase in the non-IFG-bound structures (Fig. 8a) appear to predict the conformational change observed experimentally with IFG-bound GCase. In terms of therapeutics, these data suggest that a larger molecule that mimics not only the substrate intermediate but can bind to surrounding hydrophobic subsites may be a suitable chaperone.

Although the catalytic region contains the highest number of consensus sites, the mostly highly populated consensus sites for two of the three structures is located near the N-terminus of GCase (smaller circle in Figs. 7a,
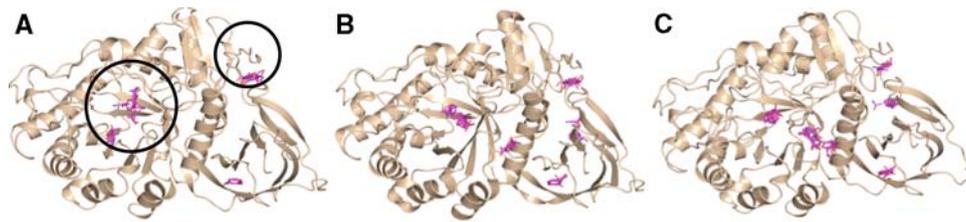
**Fig. 7** Hot spots detected on the surface of GCase using FTMap. Simulations were performed on (**a**) apo GCase (2NT1) (**b**) glycerol-bound GCase (2NSX chain A) and (**c**) IFG-bound GCase (2NSX chain B). Although differences in the location of consensus sites are evident (cluster representatives *colored* in *cyan*) two hot spot regions, *circled* in (**a**) are conserved in all three structures

**Table 4** FTMap results for glucocerebrosidase

| Region | CS ranks (size) | | | Residues |
|--------|------|------|------|----------|
| | 2nsxA | 2nsxB | 2nt1 | |
| 1 | 2(17) | 1(26) | 2(15) | D127, F128, E235, Y244, F246, Y213, W312, Y313, D315, F316, E340, C342, V343, S345, |
| | 3 (13) | 2(19) | 3(12) | E349, Q362, W381, N396, V398 |
| | 5(12) | 4(10) | 6(5) | |
| 2 | 1(20) | 3(14) | 1(20) | A1, R2, P3, C4, D24, S25, F26, R48, M49, Y418 |
| 3 | 6(10) | 6(8) | 6(6) | L67, L372, L436, V437, A438, N442, L444, D445, A446, V446 |

*CS* consensus site
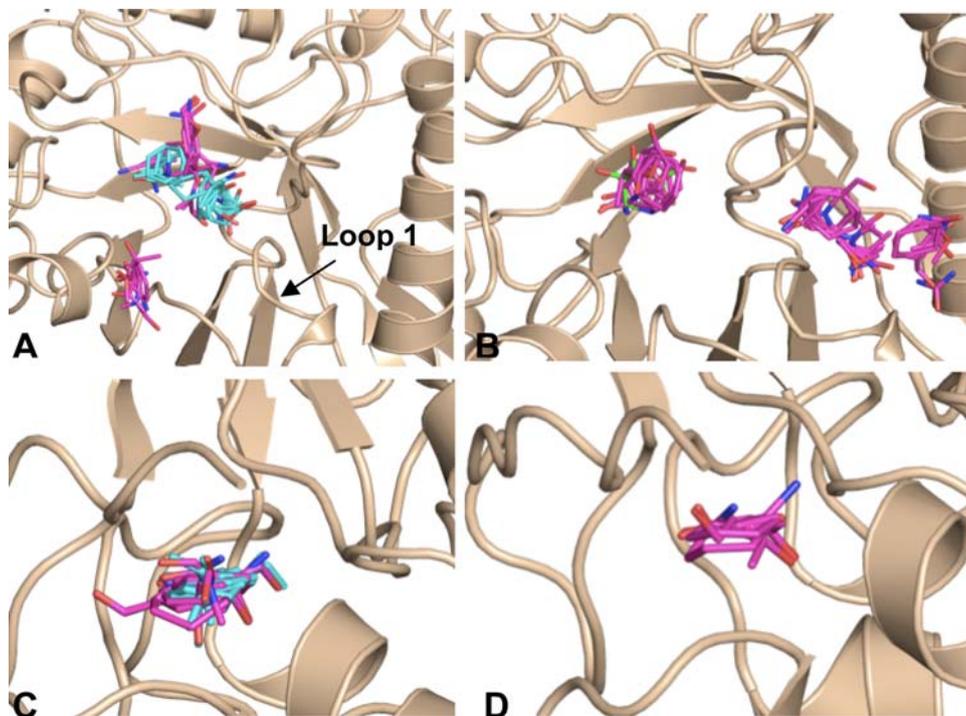


**Fig. 8** Close-up view of hot spots formed in the catalytic center (Fig. 7a, *large circle*) and N-terminal region (Fig. 7a, *small circle*) of GCase from FTMap simulations (**a**) Fragment clusters formed in catalytic center of the apo (*magenta sticks*) and glycerol-bound (*cyan sticks*) conformations of GCase. **b** Binding of IFG results in a conformational change that allows for the formation of additional hot spots in the active site region (**c**) The largest number of fragment clusters is found in the N-terminal region of GCase for both the apo (fragment clusters *colored* in *magenta*) and glycerol-bound structures of GCase (fragment clusters *colored* in *cyan*). **d** Although smaller, a hot spot is also observed in the N-terminal region of the IFM-bound structure (fragment clusters *colored* in *magenta*)

8c, d), in a groove at the interface between the β-sheet and β-barrel domains. This region coincides with cluster B identified in the MSCS experiments described above (Figs. 4b, 5b), in addition to glycerol observed previously [22]. Although not present in the apo structure (2NT1), fragment clusters are found located in a site proximal to region 2 for both glycerol- and IFG-bound GCase; based on this data, we hypothesize that a drug-sized molecule could be accommodated in this region. Overall, we observe considerable consistency of mapping data across the three structures, in terms of the size and rank of consensus sites, e.g., hot spots, determined for each region (Table 4). These new regions can now be targeted for the discovery of new remote-binding chaperones.

## Conclusions

Fragment-based methods are becoming of increased interest and application in the pharmaceutical industry. In this study we successfully applied two fragment-based approaches—MSCS and FTMap—to the identification of hot spots for DJ-1 and glucocerebrosidase, two therapeutic targets for the development of treatments for neurological disorders. Novel hot spots were identified on the surface of DJ-1 in two regions. One hot spot was found in the region containing a residue whose oxidation may protect against PD and the other in the dimer interface, where a pharmacological chaperone could be bound to increase the stability of the dimeric structure. Three regions of interest were identified for glucocerebrosidase, where multiple hot spots emerge in the catalytic region. While the catalytic function of GCase is known, interactions with substrates and other proteins are poorly understood on a structural level. Results of the MSCS experiments and FTMap simulations corroborate the location of the catalytic center, and substantiate claims of hydrophobic binding sites for the substrate ceramide. Additional hot spots found on the surface of GCase provide new hypotheses for the binding sites of key players in trafficking and catalysis, both of which are disrupted in patients with Gaucher's Disease. Taken all together, these hot spots provide novel starting points for drug discovery efforts.

In addition to these findings, excellent agreement is observed between hot spots derived from the two methods, supporting the use of computational simulations in conjunction with or in lieu of expensive and time-consuming experiments. While MSCS is a powerful approach, structural resolution is often too low to accurately determine solvent binding positions, in addition to the method itself being costly and difficult to perform. As illustrated by this study, higher resolution structures will be required to more confidently model organic heteroatoms in the GCase structure, a goal that will continue to be challenged by the somewhat lipophilic nature of GCase and subsequent tendency of cross-linked GCase crystals to disintegrate when exposed to organic solvents. In addition to successfully duplicating MSCS data using FTMap, we were able to uncover hot spots not observed in the MSCS experiments, in addition to being able to asses the binding of a much larger library of solvent-like fragments than is possible by MSCS. As is the case with rigid-body docking algorithms in general, one caveat of using FTMap is the conformational dependence of the findings; however, as illustrated by the analysis of GCase, consensus information gained from the use of multiple structures as input for FTMap simulations increases the robustness of the results. FTMap is available free to the academic community at http://FTMap.bu.edu/~FTMap/.

## References

1. Bembenek SD, Tounge BA, Reynolds CH (2009) Drug Discov Today 14:278–283. doi:10.1016/j.drudis.2008.11.007
2. Ciulli A, Williams G et al (2006) J Med Chem 49:4992–5000. doi:10.1021/jm060490r
3. Parkinson J (1817) An essay on the shaking palsy. Whitingham and Rowland, London
4. Jankovic J (2008) J Neurol Neurosurg Psychiatry 79:368–376. doi:10.1136/jnnp.2007.131045
5. Bonifati V (2007) Parkinsonism Relat Disord 13(Suppl 3):S233–S241. doi:10.1016/S1353-8020(08)70008-7
6. da Costa CA (2007) Curr Mol Med 7:650–657. doi:10.2174/156652407782564426
7. Biskup S, Gerlach M et al (2008) J Neurol 255(Suppl 5):8–17. doi:10.1007/s00415-008-5005-2
8. Wilson MA, Collins JL et al (2003) Proc Natl Acad Sci USA 100:9256–9261. doi:10.1073/pnas.1133288100
9. Moore DJ, Zhang L et al (2003) J Neurochem 87:1558–1567
10. Miller DW, Ahmad R et al (2003) J Biol Chem 278:36588–36595. doi:10.1074/jbc.M304272200
11. Rohrbach M, Clarke JT (2007) Drugs 67:2697–2716. doi:10.2165/00003495-200767180-00005

12. Liou B, Kazimierczuk A et al (2006) J Biol Chem 281:4242–4253. doi:10.1074/jbc.M511110200
13. Schmitz M, Alfalah M et al (2005) Int J Biochem Cell Biol 37:2310–2320. doi:10.1016/j.biocel.2005.05.008
14. Grace ME, Newman KM et al (1994) J Biol Chem 269:2283–2291
15. Yu Z, Sawkar AR, Kelly JW (2007) FEBS 274:4944–4950. doi:10.1111/j.1742-4658.2007.06042.x
16. Steet RA, Chung S et al (2006) Proc Natl Acad Sci USA 103:13813–13818. doi:10.1073/pnas.0605928103
17. Sawkar AR, Zimmer KD et al (2006) ACS Chem Biol 1:235–251. doi:10.1021/cb600187q
18. Compain P, Martin OR et al (2006) ChemBioChem 7:1356–1359. doi:10.1002/cbic.200600217
19. Sawkar AR, D'Haeze W, Kelley JW (2006) Cell Mol Life Sci 63:1179–1192. doi:10.1007/s00018-005-5437-0
20. Sawkar AR, Adamski-Werner SL et al (2005) Chem Biol 12:1235–1244. doi:10.1016/j.chembiol.2005.09.007
21. Sawkar AR, Cheng WC et al (2002) Proc Natl Acad Sci USA 99:15428–15433. doi:10.1073/pnas.192582899
22. Lieberman RL, Wustman BA et al (2007) Nat Chem Biol 3:101–107. doi:10.1038/nchembio850
23. Allen KN, Bellamacina CR et al (1996) J Phys Chem 100:2605–2611. doi:10.1021/jp952516o
24. Mattos C, Bellamacina CR et al (2006) J Mol Biol 357:1471–1482. doi:10.1016/j.jmb.2006.01.039
25. Mattos C, Ringe D (1996) Nat Biotechnol 14:595–599. doi:10.1038/nbt0596-595
26. Brenke R, Kozakov D et al (2009) Bioinformatics 25:621–627. doi:10.1093/bioinformatics/btp036
27. Dennis S, Kortvelyesi T, Vajda S (2002) Proc Natl Acad Sci USA 99:4290–4295. doi:10.1073/pnas.062398499
28. Kortvelyesi T, Dennis S et al (2003) Proteins 51:340–351. doi:10.1002/prot.10287
29. Silberstein M, Dennis S et al (2003) J Mol Biol 332:1095–1113. doi:10.1016/j.jmb.2003.08.019
30. Goodford PJ (1985) J Med Chem 28:849–857. doi:10.1021/jm00145a002
31. Stultz CM, Karplus M (1999) Proteins 37:512–529. doi:10.1002/(SICI)1097-0134(19991201)37:4<512::AID-PROT3>3.0.CO;2-O
32. Landon MR, Lancia DR et al (2007) J Med Chem 50:1231–1240. doi:10.1021/jm061134b
33. Otwinowski Z, Minor W (1997) Methods Enzymol Macromol Crystallogr Pt A 276:307–326
34. Vagin A, Teplyakov A (1997) J Appl Cryst 30:1022–1025. doi:10.1107/S0021889897006766
35. Murshudov GN, Vagin AA, Dodson EJ (1997) Acta Crystallogr D Biol Crystallogr 53:240–255. doi:10.1107/S0907444996012255
36. Collaborative Computational Project, Number 4 (1994) Acta Crystallogr D Biol Crystallogr 50:760–763. doi:10.1107/S0907444994003112
37. Emsley P, Cowtan K (2004) Acta Crystallogr D Biol Crystallogr 60:2126–2132. doi:10.1107/S0907444904019158
38. Kozakov D, Brenke R et al (2006) Proteins 65:392–406. doi:10.1002/prot.21117
39. Schaefer M, Karplus M (1996) J Phys Chem 100:1578–1599. doi:10.1021/jp9521621
40. Brooks BR, Bruccoleri RE et al (1983) J Comput Chem 4:187–217. doi:10.1002/jcc.540040211
41. Kosakov D, Clodfelter K et al (2005) Biophys J 89:867–875. doi:10.1529/biophysj.104.058768
42. Ruvinsky AM, Kozintsev AV (2006) Proteins 62:202–208. doi:10.1002/prot.20673
43. Blackinton R, Lakshminarasimhan M et al (2009) J Biol Chem 284:6476–6485. doi:10.1074/jbc.M806599200
44. Delano WL (2008) The PyMol molecular graphics system. Delano Scientific, Palo Alto
45. Brumshtein B, Wormald MR et al (2006) Acta Crystallogr D Biol Crystallogr 62:1458–1465. doi:10.1107/S0907444906038303
46. Kacher Y, Brumshtein B et al (2008) Biol Chem 389:1361–1369. doi:10.1515/BC.2008.163
47. Salvioli R, Tatti M et al (2005) Biochem J 390:95–103. doi:10.1042/BJ20050325
48. Reczek D, Schwake M et al (2007) Cell 131:770–783. doi:10.1016/j.cell.2007.10.018
49. de Alba E, Weiler S, Tjandra N (2003) Biochemistry 42:14729–14740. doi:10.1021/bi0301338
50. Hawkins CA, de Alba E, Tjandra N (2005) J Mol Biol 346:1381–1392. doi:10.1016/j.jmb.2004.12.045
51. John M, Wendeler M et al (2006) Biochemistry 45:5206–5216. doi:10.1021/bi051944+