Journal of Molecular Graphics and Modelling 40 (2013) 80–90

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

AutoMap: A tool for analyzing protein–ligand recognition using multiple ligand binding modes

Mark Agostino[a,b,*], Ricardo L. Mancera[a,c], Paul A. Ramsland[a,d,e,f], Elizabeth Yuriev[b]

[a] Western Australian Biomedical Research Institute, Curtin Health Innovation Research Institute, School of Biomedical Sciences, Curtin University, GPO Box U1987, Perth, WA 6845, Australia
[b] Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC 3052, Australia
[c] School of Pharmacy, Curtin University, GPO Box U1987, Perth, WA 6845, Australia
[d] Centre for Immunology, Burnet Institute, Melbourne, VIC 3004, Australia
[e] Department of Surgery Austin Health, University of Melbourne, Heidelberg, VIC 3084, Australia
[f] Department of Immunology, Monash University, Alfred Medical Research and Education Precinct, Melbourne, VIC 3004, Australia

ARTICLE INFO

Article history:
Received 2 November 2012
Accepted 1 January 2013
Available online 8 January 2013

Keywords:
Site mapping
Hot spotting
Interaction fingerprinting
Molecular docking
Protein–ligand recognition

ABSTRACT

Prediction of the protein residues most likely to be involved in ligand recognition is of substantial value in structure-based drug design. Considering multiple ligand binding modes is of potential relevance to studying ligand recognition, but is generally ignored by currently available techniques. We have previously presented the site mapping technique, which considers multiple ligand binding modes in its analysis of protein–ligand recognition. AutoMap is a partially automated implementation of our previously developed site mapping procedure. It consists of a series of Perl scripts that utilize the output of molecular docking to generate "site maps" of a protein binding site. AutoMap determines the hydrogen bonding and van der Waals interactions taking place between a target protein and each pose of a ligand ensemble. It tallies these interactions according to the protein residues with which they occur, then normalizes the tallies and maps these to the surface of the protein. The residues involved in interactions are selected according to specific cutoffs. The procedure has been demonstrated to perform well in studying carbohydrate–protein and peptide–antibody recognition. An automated procedure to optimize cutoff selection is demonstrated to rapidly identify the appropriate cutoffs for these previously studied systems. The prediction of key ligand binding residues is compared between AutoMap using automatically optimized cutoffs, AutoMap using a previously selected cutoff, the top ranked pose from docking and the predictions supplied by FTMap. AutoMap using automatically optimized cutoffs is demonstrated to provide improved predictions, compared to other methods, in a set of immunologically relevant test cases. The automated implementation of the site mapping technique provides the opportunity for rapid optimization and deployment of the technique for investigating a broad range of protein–ligand systems.

1. Introduction

Determining the structural basis of ligand–protein recognition is crucial in structure-based drug design and functional studies. Structural information can be obtained experimentally, using X-ray crystallography and NMR spectroscopy, or by automated molecular docking. While experimental techniques provide valuable insights into ligand–protein recognition, they are often laborious to implement for the rapid determination of protein complexes with distinct ligands, particularly when studying large ligand libraries and highly flexible ligands. Automated molecular docking provides rapid access to binding modes for a large number of protein–ligand systems [1]. However, it is generally accepted that there is no perfect scoring function for molecular docking, and validation studies are usually required to determine the optimal docking algorithm and scoring function for studying a given ligand–protein recognition event [1].

An issue that plagues all of the techniques is their limited consideration of the possibility of multiple binding modes [2,3]. X-ray crystallography generally provides a single "snapshot" of the ligand binding event, while an ensemble (or ensemble average) of the potential variety of binding states is generally obtained from NMR spectroscopy. A limited number of crystallographic examples where multiple ligand binding modes have been observed with a

* Corresponding author at: Western Australian Biomedical Research Institute, Curtin Health Innovation Research Institute, School of Biomedical Sciences, Curtin University, GPO Box U1987, Perth, WA 6845, Australia. Tel.: +61 892669061.
E-mail address: mark.agostino@curtin.edu.au (M. Agostino).

1093-3263/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2013.01.001

given target protein have been reported [4–8]. Molecular docking approaches can provide multiple ligand binding modes, but in practice, these are rarely utilized by the end user in describing ligand binding. Typically, only the top ranked pose (i.e., best scoring) is considered, although some efforts in considering several distinct binding modes to improve docking accuracy have been reported [9].

In our efforts to improve the reliability of binding mode prediction from docking output, we developed the site mapping technique, which utilizes an ensemble of poses obtained from molecular docking to determine the most likely protein residues involved in ligand recognition. The site mapping technique has been demonstrated to perform well in investigating peptide–antibody binding [10,11], carbohydrate–protein recognition [12–14] and carbohydrate–peptide mimicry [15,16]. It forms an integral part of the binding mode determination protocol that we have built around our mapping techniques, which also include the ligand-based epitope and conformational mapping techniques [17].

To increase the accessibility of the site mapping approach, we present AutoMap, a partially automated implementation of the technique. AutoMap requires the target protein and a series of ligand poses obtained from molecular docking as input. The interactions taking place in each pose are determined and processed to generate hydrogen bonding and van der Waals site maps of the protein. The technique is likely to be most useful for examining cases where considering a single (i.e., top ranked) docked pose may result in the wrong prediction or is not sufficiently descriptive of ligand binding, as well as studying ligand binding to shallow binding sites. The former case is exemplified by highly flexible ligands, such as carbohydrates and peptides; the latter case is characteristic of protein–protein interactions, or any protein which utilizes a large, relatively flat surface for ligand binding, such as some lectins [13]. The automated implementation allows rapid deployment of the technique for studying any protein–ligand recognition event. We have also compared the results of AutoMap with that of another mapping method, FTMap, which uses solvent-based probes to identify potential ligand binding sites in proteins.

2. Methods

2.1. The AutoMap method

The steps of AutoMap are highlighted in Fig. 1. To start, the protein–ligand system(s) of interest are identified, and the relevant structures obtained. An ensemble of predicted ligand–protein complexes is always required. The interactions taking place in each pose of the ensemble are determined and the hydrogen bonding and van der Waals site maps are generated using the appropriate cutoffs for each interaction type (Fig. 1a). The cutoffs are obtained through the use of the optional (but highly recommended) optimization procedure, using high resolution crystal structure complexes related to the target of interest (Fig. 1b).

The site mapping technique relies on the use of structural information. Hence, the minimum input for site mapping is the three-dimensional (3D) structural coordinates of the target protein and the coordinates of an ensemble of potential ligand binding modes in complex with that target. These must be supplied as two Protein Data Bank format files (PDB) [18] – one containing the target protein and the other containing the ligand ensemble. For efficient processing of the ligand ensemble, AutoMap takes advantage of the presence of specific headings in the PDB file, specifically, COMPND and TER. When present, the COMPND heading contains the name of the molecule described by the coordinates, and is used by AutoMap in naming temporary files. While defined as part of the PDB format, the COMPND heading is not typically used by many programs when creating such files. The TER heading indicates the end of the coordinates for a given molecule, and is thus used by AutoMap to split the ligand ensemble into individual ligand structures, which is necessary for processing by LigPlot [19]. Conversion of the ligand ensemble for use with AutoMap is achieved using the Silico toolkit [20], which inserts the COMPND and TER headings in the appropriate sections of the ensemble file. The steps to achieve this are described in the AutoMap manual (Supplementary Material: S1).

AutoMap has been designed to permit the use of pose ensembles generated using a variety of molecular docking programs. We have successfully applied AutoMap to poses generated by Glide [21,22], GOLD [23], DOCK [24] and AutoDock [25]. However, any docking program which either provides pose output in a standard file format (e.g., MOL2, PDB) or is able to have its output easily converted to such a file format can be used with AutoMap. When ligands containing multiple residues are docked with AutoDock, some additional steps must be taken to ensure that these can be correctly processed by LigPlot. The atomic coordinates for ligands containing multiple residues (such as carbohydrates and peptides) are typically "shuffled" by AutoDock. The lack of residue order makes automatic identification of the starting and finishing residues, required for reliable interaction determination, quite challenging. The additional steps necessary for successful processing of AutoDock poses are described in the AutoMap manual. Briefly, these entail a number of file format conversions and the use of additional scripts which can restore the original ordering of the coordinates (Supplementary Material: S1, S6).

For input into AutoMap, the pose ensemble must be collected into one file. It can consist of poses obtained from docking just one ligand, or poses of multiple related ligands. If using multiple related ligands in one site mapping run, it is important that the ligands are appropriately related, for example, by containing a particular epitope (e.g., αGal-terminating carbohydrates [12]) or belonging to a particular structural or functional class (e.g., carbohydrate-mimetic peptides [15]). In the case of using multiple related ligands, the interaction data will be collected individually for each ligand, as well as combined for all ligands; thus, it is possible to generate "per-ligand" site maps, reflecting recognition of just one ligand, as well as a "ligand ensemble" site map, reflecting recognition of a collection of related ligands.

2.2. Determination of the interactions in each docked pose

Intermolecular interactions are determined using the LigPlot program [19]. LigPlot is convenient to use for this task as it generates not only a visual interpretation of the interactions, but also tabulates the specific details of the interactions. Recently, LigPlot has been updated and re-released as LigPlot+ [26]. This new distribution still includes the original LigPlot executables, and can thus be used with AutoMap (see Supplementary Material: S1 for details). LigPlot+ is available to academic users at no charge via a limited license. LigPlot+ determines the hydrogen bonding and van der Waals interactions in a given ligand–protein complex using specific definitions as to what constitutes a hydrogen bond and a van der Waals interaction [19]. Although hydrogen bonding and van der Waals interactions are both types of non-bonded interactions, LigPlot+ specifically refers to van der Waals interactions as "non-bonded contacts" (Table 1). A hydrogen bond is determined to be present when a hydrogen bond donor (e.g., the oxygen atom of a hydroxyl group) is within both a given distance and a given angle of a hydrogen bond acceptor (e.g., a carbonyl oxygen). A van der Waals interaction is determined to be present when a carbon or sulfur atom of one molecule is within a given distance of a carbon or sulfur atom of another molecule. AutoMap uses the default angle and distance constraints used by LigPlot+ for defining

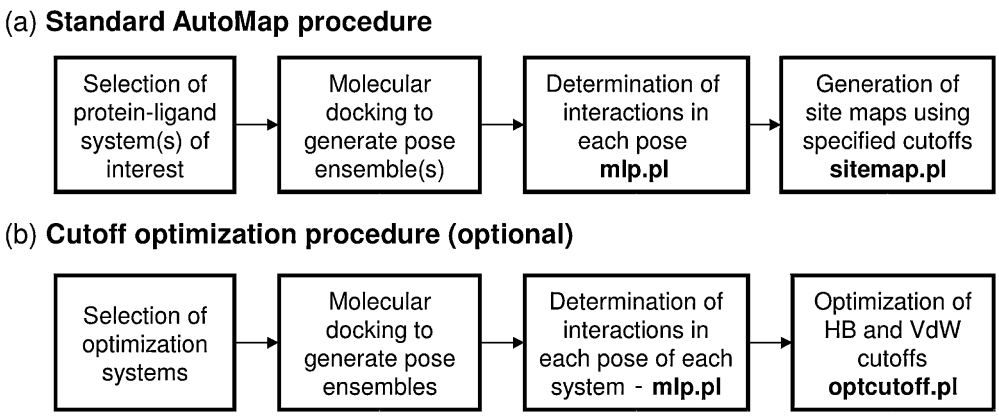

Fig. 1. Flowcharts describing the AutoMap procedure. The standard AutoMap procedure (a) and the accompanying cutoff optimization procedure (b).

intermolecular interactions. For complexes of small organic molecules with proteins, it may be necessary to cross-check the interactions determined by LigPlot+ with a direct examination of the 3D structure of the complex as the intermolecular interactions in such ligands may not be determined correctly, as has been found for GPCR ligands [27]. In these situations, it may be necessary to modify the angle and distance constraints used by LigPlot+, or to add the ligand description to the ligand dictionary. The relevant procedures are described in the LigPlot+ manual.

Interaction data for each pose generated by LigPlot+ is collected using a Perl script MLP (Supplementary Material: S2). The summary table (Table 1) generated for each complex is collected for subsequent analysis. For future reference, the user may also choose to retain the image generated by LigPlot+, the hydrogen bond table, the van der Waals interaction table and the PDB coordinates of each ligand–protein complex, as well as the summary table, by creating folders to collect this data. Since this can result in the accumulation of a large number of large files, the default mode in AutoMap is to discard these files.

2.3. Tallying of hydrogen bonding and van der Waals interactions of each pose

Tallies of the interactions observed in each pose between ligand and protein are generated with the Perl script SITEMAP (Supplementary Material: S3). Tallying follows the algorithm developed by us earlier [11,12,16]. Specifically, the interactions are tallied according to the residue with which they occur and the type of interaction, i.e., hydrogen bonding or van der Waals. Each hydrogen bond between the ligand and a specific protein residue is counted as one towards the hydrogen bond tally for that residue (Fig. 2). Due to the way in which the individual van der Waals interactions are reported by LigPlot+ (see above), directly tallying these contacts results in a significant bias towards aromatic and branched chain amino acids, which are generally capable of forming many such interactions with a specific portion of a ligand. To circumvent this inherent bias, we consider one point towards the van der Waals tally of a particular protein residue if at least one van der Waals interaction between that protein residue and a particular ligand residue is present (Table 2).

To perform site mapping on multiple ligands processed by MLP, a batch script is provided that generates "per-ligand" site maps, as well as a site map using the entire ligand ensemble (Supplementary Material: S4. MULTISITEMAP).

Table 1
Stylized version of summary table generated by LigPlot+ (ligplot.sum).[a]

	Hydrogen bonds[b]			
Interacting residues[c]	M—M	S—S	M—S	Non-bonded contacts[d]
Gal 2 → Pro107H	0	0	0	1
Gal 2 → Trp38H	0	0	0	9
Gal 2 → Trp107L	0	0	0	2
Tyr114L → Gal 1	0	1	0	0
Gal 1 → Ser108L	0	0	1	0
Lys38L → Gal 1	0	1	0	0

[a] Table summarizing the interactions in the ligand pose shown in Fig. 2. Not all van der Waals contacts are shown in Fig. 2.
[b] Types of hydrogen bonds: M—M – main chain (i.e., backbone) to main chain, S—S – sidechain to sidechain, and M—S – main chain to sidechain. All hydrogen bonds between a non-peptide ligand and a protein are treated as S—S or M—S.
[c] For hydrogen bonding interactions, the arrow points from donor to acceptor. For van der Waals interactions, the ligand residue is displayed first.
[d] LigPlot+ term for van der Waals interactions [19].

2.4. Normalization of the tallies and generation of the site maps

In addition to tallying the interactions, SITEMAP also normalizes the tallies and generates the hydrogen bonding and van der Waals site maps. The tallies are normalized by dividing the tally for a given residue by the total number of interactions observed in the entire ensemble. Thus, the normalized tallies reflect the percentage of interactions made with a particular residue against the total number of observed interactions. The normalized tallies are then used to identify the key protein residues involved in ligand recognition. Hydrogen bonding and van der Waals tallies are considered separately. For each of the hydrogen bonding and van der Waals tallies, the mapped residues are sorted from greatest to least contributing. Starting from the first entry of the sorted list, the interactions are then cumulatively summed. Once the cumulative sum reaches a given cutoff, all residues included in the sum up to this point are considered as being important for the interactions. The selected residues, along with their percentage contribution to hydrogen bonding and van der Waals interactions are given in a table as output (Table 3).

Cutoff values must be specified by the user for both hydrogen bonding and van der Waals interactions, since different cumulative sum cutoffs for hydrogen bonding and van der Waals interactions

Table 2
Example of summary table (Table 1) as processed by AutoMap.

Interacting protein residue	Hydrogen bonds	van der Waals interactions
Pro107H	0	1
Trp38H	0	1
Trp107L	0	1
Tyr114L	1	0
Ser108L	1	0
Lys38L	1	0

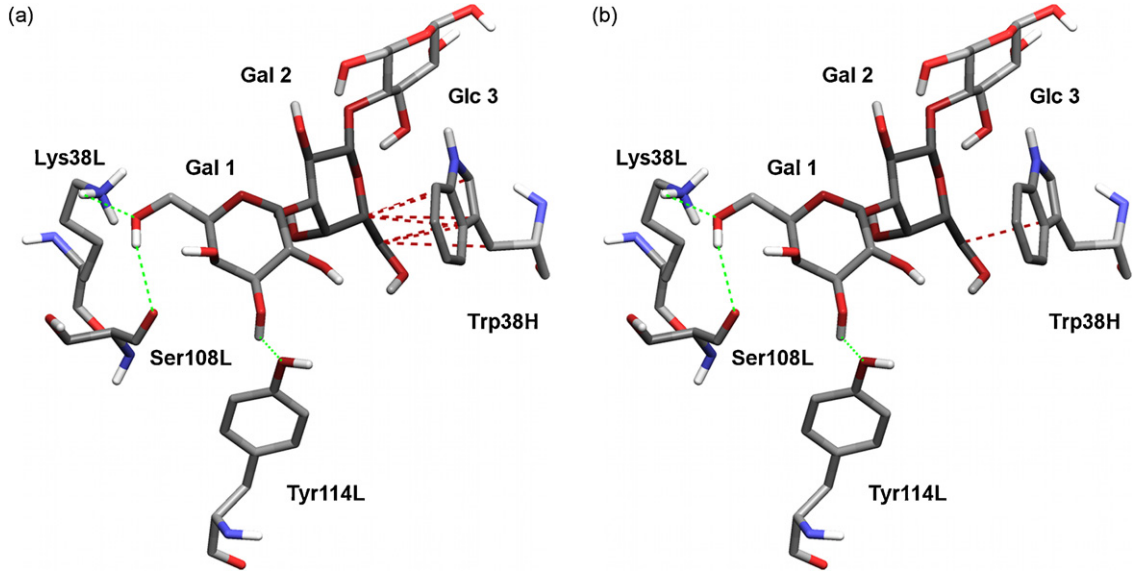

Fig. 2. Illustration of the representation of intermolecular interactions in AutoMap. The recognition of Galα(1 → 3)Galβ(1 → 4)Glcβ by the anti-αGal antibody 15.101 [17] is used as an example. All hydrogen bonds are considered individually (green dashes). Individual van der Waals interactions (red dashes) are determined by LigPlot+ as two carbon/sulfur atoms occurring below a certain distance of one another. For large residues, such as tryptophan, this results in many individual van der Waals interactions (a). For site mapping, the existence of at least one individual van der Waals interaction between a ligand residue and a protein residue is considered as one towards the van der Waals tally for that residue. Thus, the interaction between Gal 2 and Trp 38H is counted as one towards the van der Waals tally for Trp 38H (b). Figures prepared using Maestro 9.2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

may be appropriate for different systems. It is highly recommended that these values are optimized against a series of appropriate test systems (generally high resolution X-ray crystal structure complexes) using the automated optimization procedure (described below). However, a value of 80% for both types of interactions is suggested if appropriate optimization systems are not available [12,13,15].

In addition to the tabular output, PDB files that can be rendered to give a visual interpretation of the data (i.e., the "site maps") can be optionally generated by SITEMAP. Since many popular molecular graphics packages allow colouring according to the contents of the B factor column, this feature is exploited here to allow rendering of the hydrogen bonding and van der Waals site maps. More specifically, atoms with B factors from 5.0 to 37.0 are rendered by most software according to a blue-to-white ramp, and those from 37.0 to 73.0 according to a white-to-red ramp. This functionality is therefore used in the conversion of the normalized tallies to "B factors" for rendering (Fig. 3). Residues which account for 5% or fewer interactions are assigned a B factor of 37.0 (i.e., rendered white). Residues which account for 20% or greater hydrogen bonds are assigned a B factor of 73.0 (i.e., rendered red). Residues which account for 20% or greater van der Waals interactions are assigned a B factor of 5.0 (i.e., rendered blue). For residues that contribute 5-20% of the total hydrogen bonding or van der Waals interactions, the B factors are calculated according to the following expressions, where $\bar{n}$ is the normalized tally for a given residue in percentage terms:

$$B_{HB} = \frac{12 \times \bar{n}}{5} + 25$$

$$B_{VDW} = \frac{-32 \times \bar{n}}{15} + \frac{143}{3}$$

Table 3
Stylized version of SITEMAP output (site.map).[a]

Residue[b]	% of total recorded hydrogen bonds[c]	% of total recorded van der Waals interactions[c]
Arg52H	15.0	7.9
Glu100AH	12.1	3.4
Arg30CL	11.6	10.8
Arg95L	11.6	0.9
Tyr33H	11.2	10.8
Ser91L	7.6	1.3
His96H	5.7	16.2
Tyr92L	5.6	12.4
Asn52DH	5.5	3.4
Asn30AL	3.6	2.0
Leu94L	3.5	3.6

[a] Table corresponds with site maps shown in Fig. 3.
[b] Residue names presented in fixed-width manner in site.map.
[c] Decimal places are not rounded in site.map.

2.5. Automated optimization of site mapping cutoff

We have previously investigated the use of site mapping in a variety of carbohydrate–protein recognition scenarios [12,13] as well as carbohydrate–peptide mimicry [15]. Identical cumulative sum cutoffs of 80% for hydrogen bonding and van der Waals interactions appeared optimal for the majority of the test cases. However, when studying the recognition of acidic sugars by antibodies, non-identical cumulative sum cutoffs for each interaction type gave rise to better results than identical cutoffs. A 90% cutoff for hydrogen bonding and a 40% cutoff for van der Waals interactions gave rise to the optimal prediction of interacting residues, as indicated by comparison to the evaluation crystal structure complexes [28]. Since the quality of site mapping may vary depending on the ligand–protein system being examined and the selection of appropriate cumulative sum cutoffs, we have developed a procedure that can automatically identify the optimal cutoffs. Prior to describing the procedure itself, it is worth discussing the metrics used to assess the quality of generated site maps, which are used by the automated procedure.

In order to assess the quality of site mapping, the metrics of recall and precision are used. These metrics rely on knowledge of the "true" ligand contacts; that is, the contacts taking place between the ligand and the target protein in a crystal structure complex.

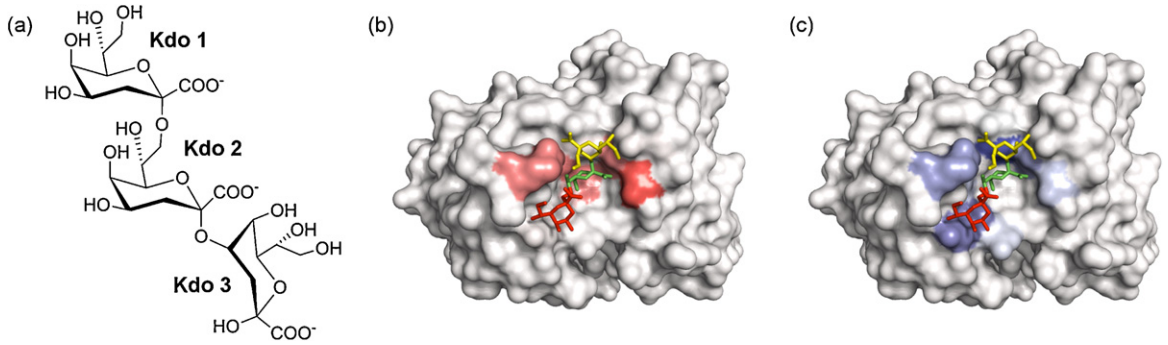

Fig. 3. Example of rendered site maps. Recognition of Kdoα(2 → 8)Kdoα(2 → 4)Kdo (a) by the antibody S25-2 (PDB 1Q9Q), as revealed by AutoMap. The hydrogen bonding (b) and van der Waals (c) site maps for S25-2 were rendered using PyMOL. The residues most intensely coloured are those predicted to be heavily involved in ligand recognition. The co-crystallized ligand is shown in each site map with the following colour scheme; Kdo 1 – yellow, Kdo 2 – green, and Kdo 3 – red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Recall (r) is the fraction of the total number of true contacts that are correctly identified by site mapping (true positives, or TP) against the total number of true contacts (the sum of true positives and false negatives, or $TP+FN$). A recall value of one indicates that all of the true contacting residues are identified by the site mapping procedure. A value of 0.5 indicates that only half of the true contacts are identified by site mapping. Precision (p) is the fraction of the total number of true contacts identified by site mapping against the total number of contacts identified by site mapping (the sum of true positives and false positives, or $TP+FP$). The precision provides an assessment of the false positive rate of site mapping. Precision values close to one indicate few false positives, while precision values close to zero indicate many false positives. In our earlier studies, the metrics of recall and precision were referred to as "reproduction" and "correctness" respectively [12,13,15,28].

In mathematical terms, recall and precision are expressed using the following formulae:

$$r = \frac{TP}{TP + FN}$$

$$p = \frac{TP}{TP + FP}$$

The F_1 score is used to provide a single value metric for quality assessment. The F_1 score is computed as the harmonic mean of recall and precision:

$$F_1 = 2 \times \frac{r \times p}{r + p}$$

The F_1 score provides a means of selecting cutoffs that provide optimal recall and precision. It reduces the likelihood of selecting low or high cutoff values, as low cutoffs (<30%) generally give high precision but low recall, and high cutoffs (>80%) generally give high recall but low precision.

The cutoff optimization procedure (Fig. 1b) requires a series of high resolution protein–ligand crystal structures related to the ultimate system of interest. Prior to carrying out cutoff optimization, cognate molecular docking and interaction determination of the docking ensemble using MLP must be performed. In addition, the list of true contacts must be supplied for each system, so that recall and precision values (and hence, F_1 scores) can be calculated. Automated cutoff optimization is performed by another Perl script (Supplementary Material: S5. OPTCUTOFF), which exhaustively generates site maps using combinations of hydrogen bonding and van der Waals cutoffs at specified intervals (i) from 0 to 100%. This script relies on the use of SITEMAP. The total number of site maps generated for each protein–ligand system (N) is determined using the following expression:

$$N = \left(\frac{100}{i} + 1\right)^2$$

For example, a 10% interval (the default used) will yield 121 site maps for each protein–ligand system, with various combinations of hydrogen bonding and van der Waals cutoffs. The total number of site maps generated for a complete validation set will therefore be equivalent to N multiplied by the number of systems in the set. The F_1 scores of reproduction and correctness for each of these site maps are collected into a table for each protein–ligand system. The values at each location of each table for each protein–ligand system are then averaged to generate a single table of averaged F_1 scores. The averaged table is then read in the direction of increasing hydrogen-bonding cutoff, followed by increasing van der Waals cutoff. The earliest instance of a maximum value in the averaged table is indicative of the optimal hydrogen bonding and van der Waals cutoffs. Thus, in cases where the same maximum F_1 score is observed in multiple places of the averaged F_1 score table, the earliest instance of that maximum is always used (i.e., lower cutoffs are generally preferred). Fig. 4 provides an illustration of the approach. The automated optimization, as implemented in OPTCUTOFF, requires approximately three minutes on an Intel Core i7 at 2.67 GHz to determine these values, representing a dramatic improvement in efficiency compared to performing a similar analysis of the data manually [12].

2.6. Comparison of the performance of automatically optimized cutoffs to other approaches

To demonstrate the utility of the automatically optimized cutoffs in site mapping, the approach was evaluated against the series of test cases that we have previously used to validate site mapping: high resolution X-ray crystal structures of carbohydrate–antibody complexes [12,29], carbohydrate–lectin complexes [13], anti-Kdo antibodies in complex with poly-Kdo carbohydrates (a model system for ganglioside–antibody recognition) [28] and peptide–antibody complexes [15]. The full list of test cases is described in Supplementary Information (S7). The docking ensembles obtained from the earlier published validation studies for each of these cases were used and the production of these is described previously [13,15,28,29]. In the case of carbohydrate– and peptide–antibody complexes, docking ensembles from Glide were used, while for carbohydrate–lectin complexes and Kdo/anti-Kdo complexes, ensembles from GOLD were used. The results obtained for site mapping using automatically optimized cutoffs for

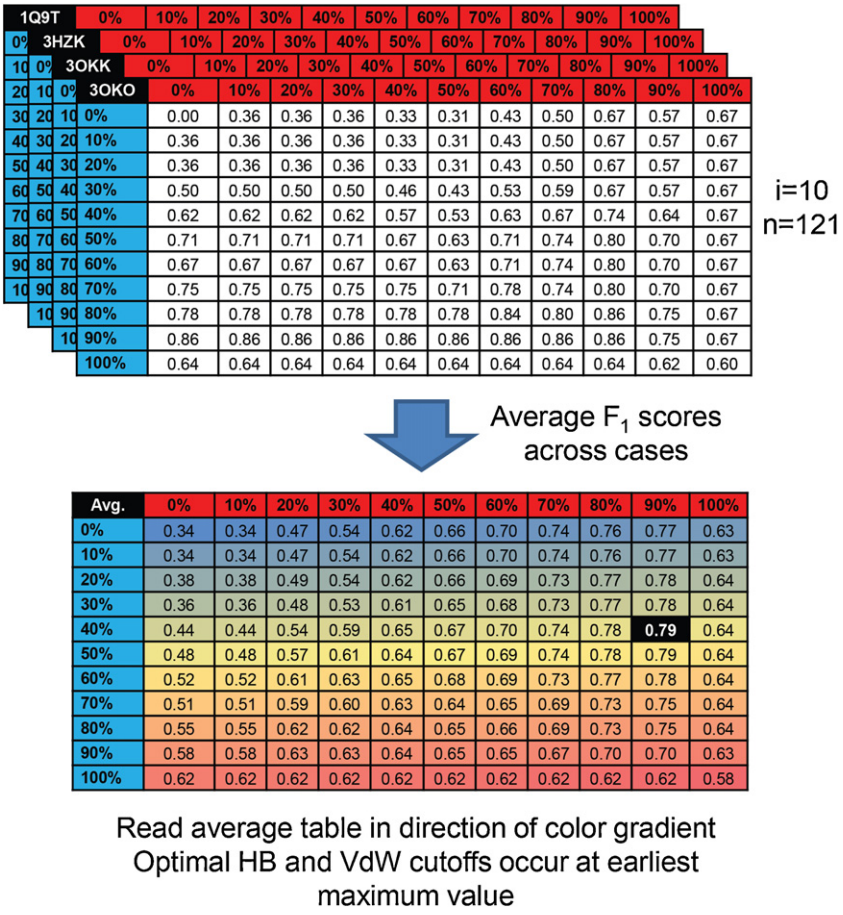

Fig. 4. Illustration of site mapping optimization. Visual representation of the procedure carried out by OPTCUTOFF. The maximum value for the averaged F_1 score is highlighted on the average table.

these test cases were compared to the results obtained using several alternative strategies. These included the use of the initially suggested cutoffs of 80% for both hydrogen bonding and van der Waals interactions, the use of the interacting residues identified by the top ranked pose, and the residues identified for importance by an alternative mapping approach, FTMap [30]. FTMap uses computational solvent mapping to identify likely binding "hot spots" on a protein surface. Solvent molecules are docked to the surface of the protein, and the hydrogen bonding and non-bonded contacts made by the docked solvent molecules are tallied for each residue. The F_1 score as described above was used to determine the success of each of these approaches in describing ligand recognition by the evaluation cases, thus also providing a means of directly comparing the performance of these approaches. In order to focus the FTMap calculations at the site of interest, all protein residues greater than 8.0 Å from the location of the bound ligand were masked (i.e., an energy penalty was applied to interactions with residues outside of the binding site). The unliganded protein and its accompanying mask were submitted to the FTMap server (http://ftmap.bu.edu) for the FTMap calculations. The lists of non-bonded and hydrogen bonding interactions indicated the residues of importance for recognition by FTMap. Residues were considered important for interactions if at least 250 hydrogen bonding interactions or at least 4500 non-bonded interactions occurred between the residue and any of the organic probes. These numbers were selected according to the lowest number of interactions received by residues involved in true contacts, averaged across the set of evaluation cases. The numbers did not vary significantly across the different sets of complexes (data not shown). The various approaches to predicting the key residues involved in ligand binding were compared by performing Wilcoxon signed rank tests in SPSS 20. In all cases, the null hypothesis was that there was no significant improvement in the F_1 scores as a result of using AutoMap with automatically optimized cutoffs as compared with using one of the other approaches. Wilcoxon tests were carried out for the entire set of test cases as one group, as well as in the context of the individual sets of test cases. The *p*-values of these tests are reported, as well as the mean, standard deviation and median for the various data sets.

3. Results

The site mapping technique has been previously demonstrated for carbohydrate–protein [12,13,28] and peptide–antibody [15] systems, and is not re-examined in detail here. When executed as described in the manual, AutoMap can readily generate site maps as in earlier reports [12,13,15,28,31,32]. Instead, the results hereby presented will focus on the use of the automated optimization procedure for identifying cutoffs for the validation systems previously studied (Supplementary Information: S7), as well as comparing the results obtained using the automatically optimized cutoffs with the initially suggested cutoffs of 80% for both hydrogen bonding and van der Waals interactions, the top ranked pose and the FTMap results. The F_1 scores obtained using the various approaches are listed in Supplementary Information (S8).

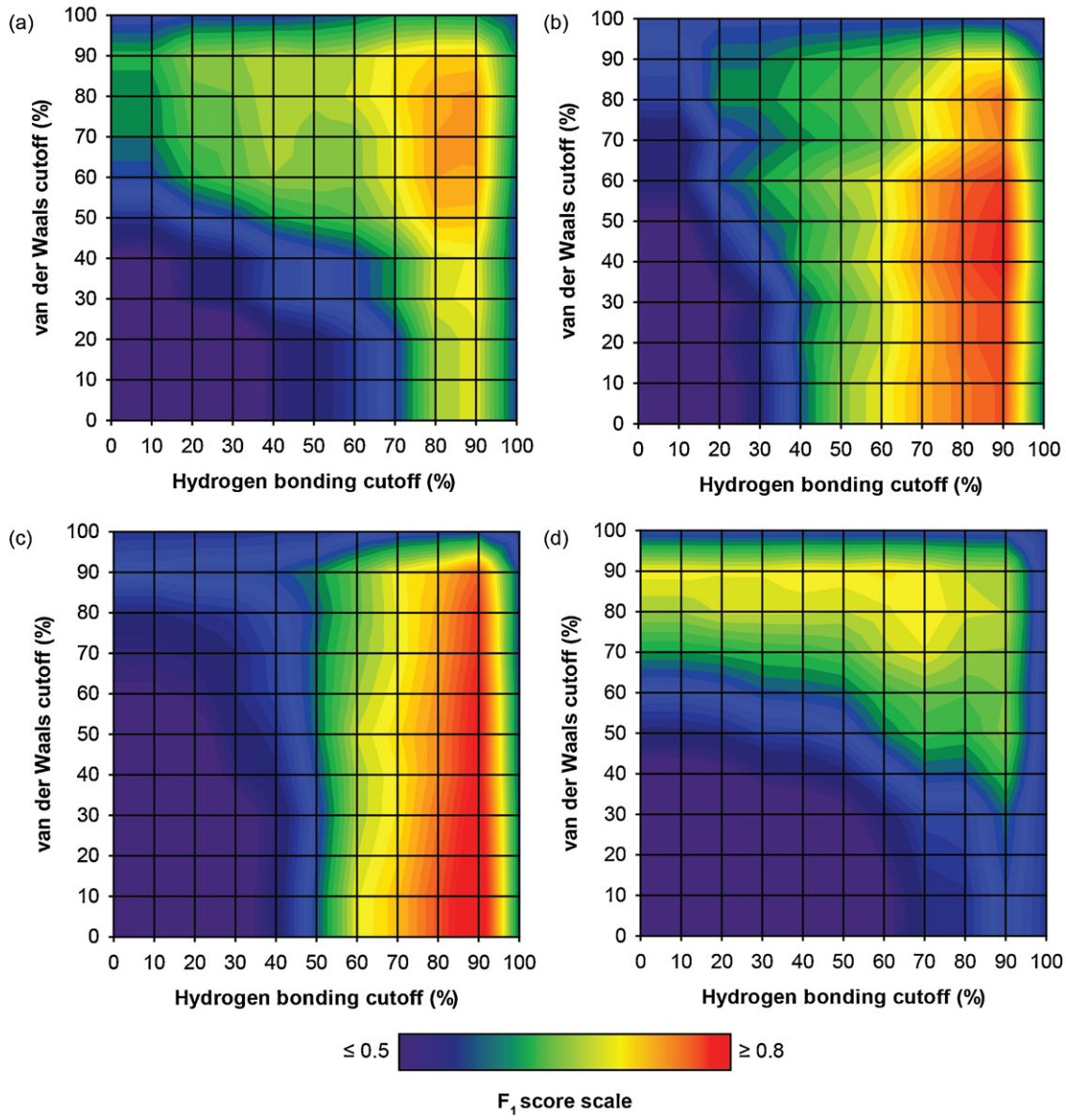

Fig. 5. Optimization of cutoffs for the sets of test cases. Plots of van der Waals cutoff vs. hydrogen bonding cutoff for the various sets of test cases, with F_1 score plotted as a coloured contour. (a) Carbohydrate–antibody recognition. (b) Ganglioside–antibody recognition. (c) Carbohydrate–lectin recognition. (d) Peptide–antibody recognition.

3.1. Optimization of cutoffs for carbohydrate–antibody recognition

The set of validation cases previously examined was used for the optimization of cutoffs for carbohydrate–antibody recognition [12]. This set of cases features high resolution structures of antibodies in complex with neutral di-, tri- and tetrasaccharides linked by $1\rightarrow2$, $1\rightarrow3$ or $1\rightarrow4$ glycosidic linkages. An optimal cutoff of 90% for hydrogen bonding interactions and 80% for van der Waals interactions was obtained (Fig. 5a). These cutoffs are similar to the cutoff of 80% for both hydrogen bonding and van der Waals interactions that were originally used (henceforce referred to as "80%:80%") [12]. The approximately equal contributions to binding made by hydrogen bonding and van der Waals interactions indicate that each interaction type is of approximately equal importance for carbohydrate–antibody recognition.

3.2. Optimization of cutoffs for ligand recognition by anti-Kdo antibodies

Ganglioside–antibody recognition is a special case of carbohydrate–antibody recognition. Gangliosides contain carbohydrate residues (specifically, sialic acids) featuring carboxylate groups and flexible chains. As there are no high resolution structures of ganglioside–antibody complexes, we have used structures of anti-chlamydial antibodies in complex with their target poly-Kdo antigens as a model for ganglioside–antibody recognition [28]. The mixed treatment of cutoffs was originally proposed for these systems, as it was apparent that a single cutoff for both hydrogen bonding and van der Waals interactions was inadequate in describing antibody recognition of gangliosides. Thus, the original analysis performed to optimize the cutoff [28] reflects a manual implementation of the cutoff optimization strategy presented here. As expected, the automated cutoff optimization

results in identical cutoffs to those obtained manually – 90% for hydrogen bonding interactions and 40% for van der Waals interactions. The plot obtained for the F_1 scores at different cutoffs (Fig. 5b) suggests that ganglioside–antibody recognition rests largely on the hydrogen bonding contribution, with the van der Waals contribution generally showing little effect on the F_1 score of the site maps. Some consideration of van der Waals interactions improves the average F_1 score of the site maps at high hydrogen bonding cutoffs (80-90%), albeit marginally.

3.3. Optimization of cutoffs for carbohydrate–lectin recognition

For carbohydrate complexes with lectins involved in innate immunity [13], the reexamination of the data identified an optimal cutoff of 90% for hydrogen bonding interactions and no consideration of van der Waals interactions. This revised cutoff scheme provides improved F_1 scores for most of the complexes compared to the previously used cutoffs, and thus results in higher quality predictions for carbohydrate–lectin recognition. The plot of F_1 score at different cutoffs (Fig. 5c) is indicative of the limited involvement of van der Waals interactions in carbohydrate–lectin recognition: essentially no variation in the F_1 score is observed for site maps with different van der Waals cutoffs but with the same hydrogen bonding cutoff. The exception to this occurs at the 100% cutoffs for both hydrogen bonding and van der Waals interactions: the F_1 score drops very quickly from 90% to 100% cutoffs. This is due to the 100% cutoffs selecting every residue with which at least one interaction was observed. This greatly reduces the precision of the site maps and hence the F_1 score is also greatly reduced. This is not unique to carbohydrate–lectin recognition, and occurs to varying extents in all of the cases examined.

3.4. Optimization of cutoffs for peptide–antibody recognition

For a set of complexes previously used to validate peptide–antibody site mapping [15], the optimal cutoffs were found to be 60% for hydrogen bonding interactions and 90% for van der Waals interactions. These cutoffs afford a slight improvement in accuracy over the originally used cutoff of 80% for both hydrogen bonding and van der Waals interactions. They also emphasize the greater importance of non-polar interactions in peptide recognition by antibodies compared to carbohydrate recognition, as previously observed [16,33,34]. This point is highlighted by the plot of F_1 score at different cutoffs for these systems (Fig. 5d), which is transposed in orientation relative to the other cases, indicating the greater power of van der Waals contacts in predicting the interacting residues. It should be noted that only a limited number of the available high resolution peptide–antibody complexes were examined here in order to make a valid comparison to the previously published study. Thus, the cutoffs presented for this type of system should be considered as approximate. Further molecular docking and site mapping studies are currently underway to optimize the cutoffs for peptide–antibody systems.

3.5. Comparison of the performance of automatically optimized cutoffs to other approaches

In applying the automatically optimized cutoffs to each of the sets of evaluation systems, a slight improvement in the quality of the site maps can be observed compared to the initially generated site maps with the 80%:80% cutoffs (Table 4). Notably, the performance is also generally improved compared to the top ranked pose. The greatest improvements relative to the top ranked pose and site maps generated with the 80%:80% cutoffs are observed for carbohydrate–lectin recognition and ganglioside–antibody recognition.

FTMap could not be successfully applied to all of the evaluation systems. The systems that failed to work generally exhibited clustering errors, suggesting that the probes failed to effectively group together in specific locations of the binding sites. It was particularly surprising to note the failure of the procedure to work on any of the anti-Kdo antibodies, which bind antigens with features perhaps the most similar to the small molecule probes utilized by FTMap. Furthermore, the procedure failed to work on some of the anti-carbohydrate antibodies with clearly defined cavities, including mAb BR96 (1CLY), mAb SYAJ/6 (1M7D) and mAb 2G12 (1OP3, 1ZLS, 1ZLU), as well as some of the anti-peptide antibodies. The protein surface mode, optimized for protein–protein interactions, also failed to provide output in any of these cases. For the remaining cases, the FTMap output was compared to the site maps generated with optimized cutoffs. In general, while FTMap can typically identify all of the residues involved in ligand recognition in the entire set of mapped residues, it failed to predict the degree of importance of the residues for ligand recognition accurately. Many residues that are not involved in ligand recognition according to experimental structures appear to have great importance according to FTMap. This greatly reduces the F_1 scores for the cases examined and hence greatly diminishes the precision of FTMap.

Wilcoxon analysis of the F_1 scores of all of the test cases as one group reveals a statistically significant improvement in using AutoMap with automatically optimized cutoffs to study ligand recognition compared to FTMap. When the sets of test cases are analyzed separately, it is revealed that both antibody and lectin recognition of carbohydrates are better modelled by AutoMap compared to FTMap. Improvement by AutoMap compared to FTMap in modelling peptide–antibody recognition is not statistically significant, while a comparison between the two methods could not be performed for ligand recognition by the anti-Kdo antibodies due to the failure of FTMap to process any of these cases. The lack of statistical significance in the set of peptide–antibody cases is most likely due to the limited number of cases that were successfully examined by FTMap.

Statistically significant improvements in the predictions made by AutoMap using the automatically optimized cutoffs over the 80%:80% cutoff were observed for the modelling of carbohydrate–lectin recognition and ligand recognition by anti-Kdo antibodies, as well as when the sets of test cases were analyzed as one group. The Wilcoxon test did not reveal a statistically significant improvement for the modelling of peptide–antibody cases; however, this could again be a consequence of the limited number of such test cases analyzed. As the automatically optimized cutoffs determined for carbohydrate–antibody recognition were similar to those selected using a manual optimization strategy, a statistically significant improvement in the optimized cutoff relative to the manually optimized cutoff was not observed. This suggests a “lucky” selection of the appropriate cutoff using the manual strategy, which failed to consider any potential differences in the contribution to recognition made by hydrogen bonds and van der Waals interactions. It should be noted that while the mean F_1 score obtained for the 80%:80% cutoff is slightly greater than that obtained for the optimized cutoffs in this case (0.76 vs. 0.74), this occurs as a result of rounding: the F_1 scores used in the Wilcoxon test were rounded to two decimal places, while the F_1 scores used throughout the automated optimization procedure were not rounded until the final report.

Statistically significant improvements in the predictions made by AutoMap using the automatically optimized cutoffs relative to the top ranked pose from docking were only observed for carbohydrate–lectin recognition. There are multiple reasons for this. Firstly, in the cases of carbohydrate–antibody recognition,

Table 4
Results of statistical analysis.[a]

	80%:80% cutoff				Top pose				FTMap				Optimized cutoffs[b]		
	p	μ	σ	Q_2	p	μ	σ	Q_2	p	μ	σ	Q_2	M	σ	Q_2
Carb:Ab	0.656	0.76	0.099	0.75	0.130	0.81	0.113	0.83	0.027	0.60	0.092	0.59	0.74	0.088	0.75
Carb:Lec	0.019	0.75	0.131	0.73	0.001	0.68	0.146	0.71	0.001	0.55	0.108	0.59	0.83	0.093	0.83
Kdo:Ab	0.035	0.73	0.054	0.73	0.176	0.85	0.109	0.88	–	–	–	–	0.80	0.069	0.80
Pep:Ab	0.149	0.69	0.094	0.69	0.310	0.68	0.114	0.69	0.109	0.55	0.111	0.54	0.71	0.088	0.69
All	0.014	0.73	0.103	0.72	0.204	0.75	0.142	0.73	0.000	0.56	0.103	0.58	0.78	0.095	0.79

[a] Mean (μ), standard deviation (σ) and median (Q_2) values specified reflect those for the F_1 score data obtained using the given method (specified in top row) for a given set of test cases (specified in left column). The F_1 score data is available in Supporting Information. The p values were obtained using the Wilcoxon signed rank test as described in the Methods.
[b] p values inapplicable for this set of data; see Section 2.

both for neutral [12] and acidic [28] carbohydrates, the top ranked poses generally match the experimentally determined structures very well. Secondly, an overall greater level of correctness/recall arises when considering the interactions in a single pose. This is because a single pose is not likely to contact a large number of residues (certainly not much larger than the number of residues contacted by the ligand in the original structure) and thus the chance of contacting non-'true' residues is overall reduced. While the Wilcoxon test failed to show a statistically significant improvement in AutoMap with automatically optimized cutoffs over the top ranked pose, such cutoffs lead to lower standard deviations in the F_1 scores relative to the top ranked poses. Thus, overall AutoMap provides more consistent prediction compared to the top ranked poses. The presence of several significant outliers in the predictions made by top ranked poses, such as in the case of LNFP-III recognition by DC-SIGN (PDB 1SL5) and Lewis Y recognition by mAb BR96 (PDB 1CLY), suggests that top poses are not reliable enough to describe recognition in every case. The lack of such outliers in the AutoMap output indicates its general greater reliability of predictions compared to the top ranked poses.

4. Discussion

Approaches which identify key residues involved in ligand binding can be loosely classified as energy-based or empirical. The well-known GRID method [35] is an example of an energy-based approach. It utilizes a water probe to map energetically favourable positions of a potential binding site. Similar methods include Multiple Copy Simultaneous Search (MCSS) [36,37] and CS-Map [38], which use a variety of small organic probes to determine favourable ligand-binding regions. PP_SITE [39] also uses small organic probes to investigate binding regions, but is targeted towards protein–protein interactions. Optimal Docking Area [40] generates a series of surface points on a protein and using increasing radii from a given surface point determines regions of negative desolvation energy. These "patches" are likely to correspond to favourable sites for protein–protein interactions. Interaction Force Fingerprints (IFFP) [41] compute the force exerted by protein residues on a bound ligand molecule, thus identifying residues exhibiting the largest contribution to ligand binding. SuperStar [42,43] is an example of an empirical approach, which uses experimental knowledge of non-bonded interactions in the crystal structures of small organic molecules to predict the likely geometry of protein–ligand interactions. Other empirical approaches incorporate the accessibility of hydrogen bonding groups [44] and the shape and extent of hydrophobic surfaces [45] to predict the importance of such regions in protein–ligand interactions [46]. Interaction fingerprint-based approaches are also examples of empirical approaches, and numerous examples of these have been reported [47,48]. The Structural Interaction Fingerprint (SIFt) method is perhaps the most well-known, and its extension, p-SIFt, can be used to illustrate the conservation of interactions between different ligands against the same target [49].

We have focused on comparing AutoMap to one of the other mapping approaches, FTMap [30]. FTMap was selected as it is under active development and readily available through a web server interface. AutoMap provided superior performance to FTMap in predicting the key binding residues of the protein. The reason for this is most likely due to the different types of probe molecules used by FTMap and AutoMap. FTMap makes use of small organic probes to identify ligand binding sites on proteins, whereas AutoMap uses the ligand itself. The way in which such probes are recognized by carbohydrate-binding proteins may not reflect the way in which carbohydrates are recognized by these proteins. Carbohydrates are not likely to be able to fill every crevice on a protein binding site, due to the generally fixed conformation of the pyranose rings of carbohydrate residues. By contrast, the probes used by FTMap are mostly solvent molecules, typically much smaller than even a single carbohydrate residue and thus more easily able to fill all of the space in the binding site. Furthermore, the types of interactions most likely to occur with carbohydrates may differ from that of the probes, which, depending on the probe used, may exhibit a bias towards a certain protein residue or interaction type. Both of these issues may be addressed by using carbohydrate-based probes with FTMap instead of its solvent-based probes. A new version of the FTMap server allows the use of custom probes (http://ftmap.bu.edu/param); however, this currently has some limitations that make both its use with carbohydrate-based ligands and its comparison with AutoMap quite challenging. The probes used must be relatively rigid, prohibiting the use of sialic acid-based probes and certain blocked variants of neutral carbohydrates (e.g., methylated derivatives of *N*-acetylglucosamine). Most importantly, the new FTMap server does not currently provide the tabular output of the original FTMap server, making a direct comparison with AutoMap, as performed here, impossible.

Despite the difficulties associated with FTMap, it is nonetheless a valuable method for identifying novel binding sites. Furthermore, AutoMap and FTMap could be employed in a complementary manner. FTMap can be used to identify binding sites, while AutoMap can be used to characterize ligand recognition at the identified sites and probe the sites in greater detail. Since the input to AutoMap is an ensemble of ligand poses obtained via molecular docking, any ligand that can be docked can also be examined by AutoMap. While AutoMap does not explicitly consider energetic contributions in generating the site maps, the ligand ensemble is based on minimizing the binding energy by the scoring function used in docking. Like the interaction fingerprint approaches, AutoMap uses interactions within given ligand–protein complexes to determine an interaction profile of the receptor. However, since AutoMap can consider multiple binding modes, the profile determined by AutoMap carries the added dimension of the importance

of a specific interaction for binding, as indicated by the normalized percentage contribution values. This assignment of the importance of a particular interaction is generally inaccessible to empirical approaches without considering some energetics. Thus, AutoMap represents an intermediate between the empirical approaches and the energy-based approaches. While energy-based studies are needed to determine the predictive accuracy of AutoMap with regard to residue/interaction importance, it nonetheless performs well for identifying the key interacting residues in a variety of systems. The promising results in the use of fingerprint approaches in ligand design [48,50,51], as well as the finding that molecular mimicry can be investigated using site mapping [15], suggest that AutoMap could be a valuable technique in structure-based drug design.

While AutoMap utilizes information about ligand–protein interaction pairs, it only provides detail of the recognition event from the point-of-view of the protein. Ligand-based mapping techniques must be used to generate the complete picture of ligand binding. We are currently optimizing our epitope and conformational mapping protocols [17] to provide a suite of computational tools for binding mode identification from docking ensembles.

5. Conclusions

We have developed AutoMap, a fully automated implementation of our previously developed site mapping procedure. AutoMap is generally applicable to studying a wide range of ligand–protein complexes, including the binding of small molecules as well as large flexible molecules such as carbohydrates and peptides. We have also developed a fully automated protocol for optimizing the cumulative sum cutoffs for use with the procedure, which exhaustively examines all possible combinations of hydrogen bonding and van der Waals cutoffs. We used this protocol to obtain improved cutoffs to those used in previously studied cases of ligand–protein recognition. Most importantly, AutoMap and its accompanying cutoff optimization procedure are very fast to run on modern computers, making it possible to use the procedure to generate site maps profiling many ligand–protein recognition events with great speed and efficiency.

The current version of AutoMap (as at the time of this writing) is provided in Supporting Information. Examples of some steps of the procedure are also described in Supporting Information (S9). The latest version of AutoMap, as well as the files needed to perform the examples, can be obtained from http://ligmap.sourceforge.net.

Authors' contributions

MA, PAR and EY conceived the ideas. MA prepared the scripts, carried out the studies and prepared the manuscript. PAR, EY and RLM edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This research was supported by a small grant from the Faculty of Pharmacy and Pharmaceutical Sciences, Monash University, to E.Y. M.A. was a recipient of the Monash University Postgraduate Publication Award (PPA). P.A.R. is the Sir Zelman Cowen Senior Research Fellow (Sir Zelman Cowen Fellowship Fund, Burnet Institute). The authors gratefully acknowledge the contribution to this work of the Victorian Operational Infrastructure Support Program received by the Burnet Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2013.01.001.

References

[1] E. Yuriev, M. Agostino, P.A. Ramsland, Challenges and advances in computational docking: 2009 in review, Journal of Molecular Recognition 24 (2011) 149–164.
[2] P.J. Lewi, M. de Jonge, F. Daeyaert, L. Foymans, M. Vinkers, J. Heeres, et al., On the detection of multiple-binding modes of ligands to proteins from biological, structural and modeling data, Journal of Computer-Aided Molecular Design 17 (2003) 129–134.
[3] B. Gorelik, A. Goldblum, High quality binding modes in docking ligands to proteins, Proteins 71 (2008) 1373–1386.
[4] K.M. Campbell, K.J. Lumb, Structurally distinct modes of recognition of the KIX domain of CBP by Jun and CREB, Biochemistry 41 (2002) 13956–13964.
[5] M.R. Wester, E.F. Johnson, C. Marques-Soares, P.M. Dansette, D. Mansuy, C.D. Stout, Structure of a substrate complex of mammalian cytochrome P4502C5 at 2.3 angstrom resolution: evidence for multiple substrate binding modes, Biochemistry 42 (2003) 6370–6379.
[6] M.D. Jacobs, J. Black, O. Futer, L. Swenson, B. Hare, M. Fleming, et al., Pim-1 ligand-bound structures reveal the mechanism of serine/threonine kinase inhibition by LY294002, Journal of Biological Chemistry 280 (2005) 13728–13734.
[7] A.S. Goré, N. Robertson, J.C. Errey, I. Ng, K. Hollenstein, B. Tehan, et al., Structure of the adenosine A_{2A} receptor in complex with ZM241385 and the xanthines XAC and caffeine, Structure 19 (2011) 1283–1293.
[8] R.E. Watkins, G.B. Wisely, L.B. Moore, J.L. Collins, M.H. Lambert, S.P. Williams, et al., The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity, Science 292 (2001) 2329–2333.
[9] P. Källblad, R.L. Mancera, N.P. Todorov, Assessment of multiple binding modes in ligand–protein docking, Journal of Medicinal Chemistry 47 (2004) 3334–3337.
[10] E. Yuriev, P.A. Ramsland, A.B. Edmunson, Docking of combinatorial peptide libraries into a broadly cross-reactive human IgM, Journal of Molecular Recognition 14 (2001) 172–184.
[11] E. Yuriev, P.A. Ramsland, A.B. Edmundson, Recognition of IgG-derived peptides by a human IgM with an unusual combining site, Scandinavian Journal of Immunology 55 (2002) 242–255.
[12] M. Agostino, M.S. Sandrin, P.E. Thompson, E. Yuriev, P.A. Ramsland, In silico analysis of antibody–carbohydrate interactions and its application to xenoreactive antibodies, Molecular Immunology 47 (2009) 233–246.
[13] M. Agostino, E. Yuriev, P.A. Ramsland, A computational approach for exploring carbohydrate recognition by lectins in innate immunity, Frontiers in Immunology 2 (2011) 23.
[14] J. Milland, E. Yuriev, P.-X. Xing, I.F.C. McKenzie, P.A. Ramsland, M.S. Sandrin, Carbohydrate residues downstream of the terminal Galα(1,3)Gal epitope modulate the specificity of xenoreactive antibodies, Immunology and Cell Biology 85 (2007) 623–632.
[15] M. Agostino, M.S. Sandrin, P.E. Thompson, P.A. Ramsland, E. Yuriev, Peptide inhibitors of xenoreactive antibodies mimic the interaction profile of the native carbohydrate antigens, Biopolymers 96 (2011) 193–206.
[16] E. Yuriev, M.S. Sandrin, P.A. Ramsland, Antibody–ligand docking: insights into peptide–carbohydrate mimicry, Molecular Simulation 34 (2008) 461–468.
[17] M. Agostino, M.S. Sandrin, P.E. Thompson, E. Yuriev, P.A. Ramsland, Identification of preferred carbohydrate binding modes in xenoreactive antibodies by combining conformational filters and binding site maps, Glycobiology 20 (2010) 724–735.
[18] Worldwide Protein Data Bank, Protein Data Bank Contents Guide: Atomic Coordinate Entry Format Description Version 3.30, 2011.
[19] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions, Protein Engineering 8 (1995) 127–134.
[20] D.K. Chalmers, B.P. Roberts, Silico – A Perl Molecular Modelling Toolkit, 2011.
[21] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Kilcic, D.T. Mainz, et al., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, Journal of Medicinal Chemistry (2004) 1739–1749.
[22] Schrödinger, LLC. Glide, Version 5.6, New York, NY, 2010.
[23] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R.D. Taylor, Improved protein–ligand docking using GOLD, Proteins 52 (2003) 609–623.
[24] P.T. Lang, S.R. Brozell, S. Mukherjee, E.F. Pettersen, E.C. Meng, V. Thomas, et al., DOCK 6: combining techniques to model RNA–small molecule complexes, RNA 15 (2009) 1219–1230.
[25] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, et al., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, Journal of Computational Chemistry 19 (1998) 1639–1662.
[26] R.A. Laskowski, M.B. Swindells, LigPlot+: multiple ligand–protein interaction diagrams for drug discovery, Journal of Chemical Information and Modeling 51 (2011) 2778–2786.

[27] F.M. McRobb, B. Capuano, I.T. Crosby, D.K. Chalmers, E. Yuriev, Homology modeling and docking evaluation of aminergic G protein-coupled receptors, Journal of Chemical Information and Modeling 50 (2010) 626–637.
[28] M. Agostino, P.A. Ramsland, E. Yuriev, Antibody recognition of cancer-related gangliosides and their mimics investigated using *in silico* site mapping, PLoS ONE 7 (2012) e35457.
[29] M. Agostino, C. Jene, T. Boyle, P.A. Ramsland, E. Yuriev, Molecular docking of carbohydrate ligands to antibodies: structural validation against crystal structures, Journal of Chemical Information and Modeling 49 (2009) 2749–2760.
[30] R. Brenke, D. Kozakov, G.-Y. Chuang, D. Beglov, D. Hall, M.R. Landon, et al., Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques, Bioinformatics 25 (2009) 621–627.
[31] M. Agostino, W. Farrugia, M.S. Sandrin, A.M. Scott, E. Yuriev, P.A. Ramsland, Structural glycobiology of antibody recognition in xenotransplantation and cancer immunotherapy, in: P. Kosma, S. Müller-Loennies (Eds.), Anticarbohydrate Antibodies – From Molecular Basis to Clinical Application, Springer, New York, 2012, pp. 203–228.
[32] M. Agostino, M.S. Sandrin, P.E. Thompson, W. Farrugia, P.A. Ramsland, E. Yuriev, Carbohydrate-mimetic peptides: structural aspects of mimicry and therapeutic implications, Expert Opinion on Biological Therapy 11 (2011) 211–224.
[33] F.-X. Theillet, F.A. Saul, B. Vuillez-Le Normand, S. Hoos, F. Felici, A. Weintraub, et al., Structural mimicry of O-antigen by a peptide revealed in a complex with an antibody raised against *Shigella flexneri* serotype 2a, Journal of Molecular Biology 388 (2009) 839–850.
[34] N.K. Vyas, M.N. Vyas, M.C. Chervenak, D.R. Bundle, B.M. Pinto, F.A. Quiocho, Structural basis of peptide–carbohydrate mimicry in an antibody-combining site, Proceedings of the National Academy of Sciences of the United States of America 100 (2003) 15023–15028.
[35] P.J. Goodford, A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, Journal of Medicinal Chemistry 28 (1985) 849–857.
[36] A. Miranker, M. Karplus, Functionality maps of binding sites: a multiple copy simultaneous search method, Proteins 11 (1991) 29–34.
[37] P.D.J. Grootenhuis, M. Karplus, Functionality map analysis of the active site cleft of human thrombin, Journal of Computer-Aided Molecular Design 10 (1996) 1–10.
[38] M.R. Landon, D.R. Lancia Jr., J. Yu, S.C. Thiel, S. Vajda, Identification of hot spots within druggable binding regions by computational solvent mapping of proteins, Journal of Medicinal Chemistry 50 (2007) 1231–1240.
[39] Y. Gao, R.X. Wang, L.H. Lai, Structure-based method for analyzing protein–protein interfaces, Journal of Molecular Modeling 10 (2004) 44–54.
[40] J. Fernandez-Recio, M. Totrov, C. Skorodumov, R. Abagyan, Optimal Docking Area: a new method for predicting protein–protein interaction sites, Proteins 58 (2005) 134–143.
[41] H. Shadnia, J.S. Wright, J.M. Anderson, Interaction force diagrams: new insight into ligand–receptor binding, Journal of Computer-Aided Molecular Design 23 (2009) 185–194.
[42] M.L. Verdonk, J.C. Cole, P. Watson, V. Gillet, P. Willett, SuperStar:, Improved knowledge-based interaction fields for protein binding sites, Journal of Molecular Biology 307 (2001) 841–859.
[43] M.L. Verdonk, J.C. Cole, R. Taylor, SuperStar: a knowledge-based approach for identifying interaction sites in proteins, Journal of Molecular Biology 289 (1999) 1093–1108.
[44] M.D. Kelly, R.L. Mancera, A new method for estimating the importance of hydrogen-bonding groups in the binding site of a protein, Journal of Computer-Aided Molecular Design 17 (2003) 401–414.
[45] M.D. Kelly, R.L. Mancera, A new method for estimating the importance of hydrophobic groups in the binding site of a protein, Journal of Medicinal Chemistry 48 (2005) 1069–1078.
[46] M.D. Kelly, R.L. Mancera, Comparative analysis of the surface interaction properties of the binding sites of CDK2, CDK4 and ERK2, ChemMedChem 1 (2006) 366–375.
[47] S.C. Brewerton, The use of protein–ligand interaction fingerprints in docking, Current Opinion in Drug Discovery and Development 11 (2008) 356–364.
[48] M.D. Kelly, R.L. Mancera, Expanded interaction fingerprint method for analyzing ligand binding modes in docking and structure-based design, Journal of Chemical Information and Computer Science 44 (2004) 1942–1951.
[49] J. Singh, Z. Deng, G. Narale, C. Chuaqui, Structural interaction fingerprints: a new approach to organizing, mining, analyzing and designing protein–small molecule complexes, Chemical Biology & Drug Design 67 (2006) 5–12.
[50] S. Sciabola, R.V. Stanton, J.E. Mills, M.M. Flocco, M. Baroni, G. Cruciani, et al., High-throughput virtual screening of proteins using GRID molecular interaction fields, Journal of Chemical Information and Modeling 50 (2010) 155–169.
[51] T. Sato, T. Honma, S. Yokoyama, Combining machine learning and pharmacophore-based interaction fingerprint for in silico screening, Journal of Chemical Information and Modeling 50 (2010) 170–185.