ABSTRACT
The design and implementation of a new algorithm, known as PROXIMO for protein oxidation interface modeller, is described to predict the structure of protein complexes using data generated in radical probe mass spectrometry (RP-MS) experiments. Photochemical radiolysis and discharge sources can be used to effect RP-MS in which hydroxyl radicals are formed directly from the bulk solvent on millisecond timescales and react with surface accessible residues in footprinting-like experiments. The algorithm utilizes a geometric surface fitting routine to predict likely structures for protein complexes. These structures are scored based on a correlation between the measured solvent accessibility of oxidizable residue side chains and oxidation shielding data obtained by RP-MS. The algorithm has been implemented to predict structures for the ribonuclease S-protein–peptide and calmodulin–melittin complexes using RP-MS data generated in this laboratory. The former is in close agreement with the high-resolution experimental structure available.

Contact: kdownard@usyd.edu.au

INTRODUCTION
High-resolution structures of protein complexes are typically solved by employing X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy. High resolution is defined in the context of X-ray studies by the smallest deviation in the diffraction pattern with d-spacings of typically <1.8 Å. Despite the strengths of these experimental techniques, they both have some limitations for studies of macromolecular complexes. Apart from being experimentally and analytically time-consuming, both approaches require relatively large amounts (milligram levels) of purified protein that often preclude the study of proteins that are expressed at low levels or which are difficult to isolate from biological sources. The study of proteins at relatively high concentrations can result in molecular aggregation that is not representative of their physiological state. X-ray crystallography suffers from the requirement that protein complexes be maintained within a crystal lattice that can disrupt the interaction and precludes the study of molecular dynamics. Improvements in NMR spectroscopy in recent years, including the use of high-field magnets and techniques such as transverse relaxation-optimized spectroscopy (TROSY), have led to the study of large macromolecules (Fernandez and Wider, 2003) and their complexes with molecular weights up to 50 kDa, but above this some compromise is made in terms of resolution. Severe line broadening in NMR spectra for large macromolecules is compounded when studying their complexes. This is attributed to a reduction in the rotational diffusion rate that decreases linearly with molecular size and results in the overlap of resonance signals. In addition, the signal intensity in an NMR spectrum decreases as the molecular weight of a macromolecule increases. Reflecting these difficulties, the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (Berman et al., 2000) contained over 33 000 structures as of October 2005, while <5% of these structures represent protein complexes.

In order to compensate for this dearth of structural data for protein complexes, mass spectrometry has been applied on a number of fronts in recent years to probe the interactions of proteins. This has included the direct detection of protein complexes by both electro-spray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) mass spectrometry (Loo, 1997; Farmer and Caprioli, 1998, Kiselar and Downard, 2000), the preservation of protein complexes on MALDI targets (Kiselar and Downard, 1999, Morrissey and Downard, 2005), and solution-based studies in which protein complexes are first labelled or enzymatically cleaved at sites within the complex that are accessible to the chemical agent or enzyme. The labelled or proteolysis products are then analysed by mass spectrometry to identify the sites of modification or cleavage.

One of the most common of such methods involves hydrogen–deuterium exchange (Engen and Smith, 2001), the level or rate of exchange at the amide hydrogens across the backbone is measured based upon a shift in the mass of the proteolytic peptides. When results for free and complexed proteins are compared, those residues within the interaction interface can be identified by marked reductions in the exchange of amide hydrogen with the deuterated solvent. The approach suffers from difficulties with preventing back-exchange prior to or during analysis of the proteolytic peptides.

Cross-linking experiments, in which residue side chains that are in close proximity are covalently linked by a chemical reagent, provides another means by which the interfaces between complexed proteins can be examined (Bennett et al., 2000). However, there remains an ongoing concern that the cross-linking agents themselves may disrupt or perturb the structure of the protein complex in these experiments, and some of the chemistries employed during coupling of the cross-linker can be adverse to the preservation of protein structures (Peters and Richards, 1997).

As first reported in 1999 (Maleknia et al., 1999a), we have developed a new experimental approach known as radical probe mass spectrometry (RP-MS) with which to probe the structure of...
proteins and their interaction with other macromolecules. In RP-MS, proteins or their complexes are treated with high fluxes of the hydroxyl radical on millisecond timescales. This results in the limited oxidation of amino acid side chains at residues exposed to the solvent (Maleknia et al., 1999b). The small size of the probe and the irreversible chemistry ensures the approach can monitor subtle differences in a protein’s structure or the level of shielding within a protein–protein interface. A significant body of published work has shown that, in the case of exposures on short millisecond timescales, there is no measurable change or damage to a protein’s structure, and protein complexes are maintained (Maleknia et al., 2001, Maleknia and Downard, 2001, Maleknia et al., 2004, Wong et al., 2003, 2005). On longer timescales (>50 ms), protein degradation and cross-linking has been observed (Maleknia and Downard, 2001) and can be exploited to biophysically study the onset of such damage in the context of disease and aging (Shum, 2005). It is important to note that in RP-MS (Maleknia et al., 1999a,b) the hydroxyl radical is generated from the bulk solvent and not from hydrogen peroxide or other chemical agents added to solution that can be deleterious in maintaining correct protein conformations and can result in direct oxidation pathways from the oxidizing agent itself over the hydroxyl radical.

The nature of a protein interaction interface is examined by exposing the protein complex to radicals under two separate solution conditions, one at which the complex is preserved and a second where it is dissociated, often by varying the pH of the solution (Wong et al., 2003, 2005). Differences in the levels of oxidation at reactive residues within the interaction interface are observed. These are quantified to produce a measure of the bulk solvent accessibility at each of the reactive residue side chains. From the extent of shielding, it is possible to construct a three-dimensional representation of the protein–protein interface with single amino acid resolution at the reactive residues.

To enable structures for protein complexes to be proposed from their component molecules and the RP-MS data, a new docking algorithm has been developed and implemented. Since the development of the first docking program DOCK (Kuntz et al., 1982; Ewing et al., 2001), application of docking strategies in structural biology and for rational drug design have significantly advanced (Goldman and Wipke, 2000; Halperin et al., 2002; Chen and Weng, 2003; van Dijk et al., 2005). One data-driven docking algorithm is known as HADDOCK (High Ambiguity Driven DOCKing) (Dominguez et al., 2003). HADDOCK utilizes ambiguous interaction restraints (AIRs) and data obtained in NMR spectroscopy or mutagenesis experiments. Residues are defined as passive or active in the context of their significance in complex formation. Passive residues must have side chains that are at least 50% accessible to the bulk solvent under this definition.

The algorithm described in this article, known as PROMIXO for PRotein OXidation Interface Modeller, employs quantitated oxidation levels measured at reactive residues in RP-MS experiments. Differences in the levels of oxidation at reactive residues in the proteins alone and in complex, together with a geometric fitting routine, are used to assemble and score the structures for proposed protein complexes. Importantly, the algorithm employs an intuitive graphical user interface (GUI) to aid data entry and analysis of the results. This article describes the design and implementation of the PROXIMO algorithm and its application to the study of two protein complexes previously probed by RP-MS. The results are discussed in the context of experimentally obtained and theoretically modelled structures for the protein complexes.

**SYSTEMS AND METHODS**

The PROXIMO algorithm was written in the ANSI (American National Standards Institute, 1998) C++ programming language. Source code libraries, particularly the Biochemical Algorithm Library (BALL) (Kohlbacher and Lenhof, 2000), were used to develop PROXIMO for the analysis of molecular coordinate data (especially PDB files) and docking. The Boost (Gregor, 2005, http://www.boost.org/) serialization library was used to implement the saving of output results. QT (release 3.3.4), a cross platform C++ GUI library, was used to develop the GUI for PROXIMO. The KDEvelop (version 3.2.2) integrated development environment (IDE) was used in combination with QTDesigner (version 3.3.4) to construct the GUI. KDEvelop was configured to use the GNU C++ (version 4.0.1) compiler.

All development and testing of PROXIMO was performed using two personal computers. One featured an AMD Athlon 64, 3 GHz processor with 1 GB of random access memory (RAM) while the other featured an Intel Centrino 1.7 GHz processor with 512 Mb of RAM. Both the systems ran the open source Red Hat sponsored Fedora Core 4 Linux distribution (Kernel version 2.6.12-1).

The PROXIMO source consists of ~4000 lines of code. The algorithm reads and saves the PDB files and calculates the solvent accessibility surface (SAS) within seconds depending on the size of the molecule being analyzed. The algorithm performs the rigid body docking step in ~60 min when the default values are used. Scoring of the 2000 highest-ranked complexes is achieved in ~5 min dependent upon the size of the molecules and the amount of RP-MS data. The time required for clustering and root mean square deviation (RMSD) calculations are highly dependent on molecular size. PROXIMO was coded with multi-threading support to enable navigation of the GUI and other use of the computer while the calculations are performed. However, owing to the intensive use of the processor during docking, it is recommended that the computer that runs PROXIMO not be used simultaneously to perform other computationally demanding procedures. Although testing was performed with a Linux operating system, all components and libraries are platform independent that enables the algorithm to be used on current Microsoft Windows and Macintosh OSX operating systems.

**ALGORITHM**

An overview of the PROXIMO algorithm is represented schematically in Figure 1. The user enters the PDB files representing the structures of each of the molecules in complex. The nine reactive residues (Met, Cys, Trp, Tyr, Phe, His, Pro, Leu and Lys) in RP-MS (Maleknia et al., 1999b) are identified by the algorithm and SAS calculated at the side chains of these residues. The user then inputs the oxidation shielding data at each of the reactive residues. This is measured in RP-MS experiments as the total area under selected ion chromatograms (SIC) for all ions of each oxidized peptide divided by the total area for total peptide (oxidized and unoxidized) and expressed as a percentage (Wong et al., 2005). Where oxidation is not measured at a reactive residue, the field is left blank.

The co-ordinates for the designated mobile molecule are then rotated about the static molecule in a geometric fitting routine. In this regard, PROXIMO employs a Katchalski-Katzir geometric fitting routine (Katchalski-Katzir et al., 1992). In this routine, a mathematical representation of the structure of each of the proteins is obtained by converting their atomic coordinates into two three-dimensional grids or matrices a and b. Each grid point is then assigned a value depending on whether it represents a point inside, outside or on the surface of the molecule according to...
where $x, y$ and $z$ are the indices of the three-dimensional matrix and $\rho$ and $\sigma$ have uncommon values of either $-15$ or $1$ that are used in scoring to establish surface overlap and penetration. Grid cells within a distance of $r$ from at least one atom nucleus are deemed to be inside the molecule while those within a thin surface layer $s$ are distinguished as being on the surface of the molecule. The geometric fit of the proteins in complex is evaluated by keeping one protein defined by matrix $a$ static while rotating the one represented by matrix $b$ around it and calculating a correlation value $c$ by Equation (1).

$$c_{\alpha, \beta, \gamma} = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} a_{x, y, z} \cdot b_{x+i, y+j, z+k}$$

Figure 1. Schematic representation of the PROXIMO algorithm showing each of the computational steps.
Iteration through all values of $\alpha, \beta$ (from 0 to 360°, in 12° steps as the default) and $\gamma$ (0–180°, in 12° steps as the default) allows for all conformations for the protein complex to be searched. Since such a calculation requires $N^3$ multiplications and additions for each of the $N^3$ relative steps for values $(\alpha, \beta, \gamma)$, the overall computation requires in the order of $N^6$ calculations. In order to avoid this lengthy calculation, the Katchalski-Katzir algorithm utilizes a fast Fourier transformation (FFT) to compute the correlation more rapidly (Katchalski-Katzir et al., 1992).

The PROXIMO algorithm then utilizes the data generated in RP-MS experiments to score the hypothetical complexes. The level of oxidation at reactive residues measured in RP-MS experiments has been shown to be proportional to the SAS of the residue side chain (Kisela et al., 2002; Malekina and Downard, 2001). Therefore, for protein complexes, it is possible to identify those residues that are shielded from the solvent, and the level of shielding that occurs upon the interaction of the proteins that bind, by comparing the extent of oxidation at reactive residues when the proteins are in their free and complexed form. Ideally, the level of oxidation (expressed as a percentage) should be determined at each individual reactive residue following proteolysis of the proteins in both their free and complexed forms. This is achievable if the proteolytic digest produces peptides containing one reactive residue per peptide. In order to obtain such peptides, several different proteases can be used alone or in combination. The site of oxidation can be confirmed by tandem mass spectrometry. If the resultant peptides still contain several reactive residues, tandem mass spectrometry can be used to unambiguously identify the sites of oxidation, though this is a less reliable method of calculating the levels at each reactive residue since the energetics of ion fragmentation influence product ion yields. In this case, common oxidation levels are entered for all reactive residues. It will be shown in the subsequent applications of PROXIMO that structures in close accord with X-ray crystallographic data can still be obtained under these circumstances since the geometric fitting routine imposes constraints on the number of possible structures.

Conformations generated by the geometric fit algorithm are scored in PROXIMO based upon the correlation between oxidation shielding and decrease in the SAS value at each reactive residue. Calculation of the SAS value involves defining a Connolly (Connolly, 1983a,b) surface for the molecule by computationally rolling a sphere with radius equal to that of the oxidising agent (1.0 Å for the hydroxyl radical) over the atomic representation. From the resultant surface representation, the area of the side chains is calculated in units of Å². In order to score a hypothetical complex, the SAS of reactive residues of the complexes generated by the geometric fitting routine are also calculated. By comparing the SAS in the free form with that observed in the complex form, a percentage shielding for each residue is obtained.

Once SAS shielding values have been calculated, these values and the user input oxidation shielding are used in calculating a score for each of the hypothetical complexes using Equation (2):

$$\text{score} = \frac{\sum_{r=1}^{N} |O(r, N_r, c) - O(r, S(r))|}{N_r} + m \frac{\sum_{c=1}^{N} |S(r) - S(r)|}{N_c}$$

where $N_r$ is the total number of reactive residues for which oxidation has been observed by RP-MS and the subscripts $s$ and $c$ denote those residues for which oxidation has been observed by RP-MS and the subscripts $s$ and $c$ denote those residues that are shielded and not shielded (or constant), $r$ is a reactive residue, $O(r)$ is the oxidation shielding of residue $r$, and $S(r)$ is the SAS shielding of $r$. The oxidation shielding is calculated from the difference in oxidation levels before and after complexation divided by the oxidation level modifier $m$. This is used in order to adjust the impact of shielded residues compared with constant residues on the final score. Essentially, Equation (2) is a method of determining the difference between experimentally determined shielding from oxidation and the SAS shielding calculated upon formation of a protein complex. The difference in shielding $|O(r) - S(r)|$ is multiplied by the oxidation shielding in order to adjust the score to the magnitude of the experimental results. For example, if the $O(r)$ and $S(r)$ of residue $r$ were 1 and 11% respectively, the absolute difference between them is 10%. If $O(r)$ and $S(r)$ were 50 and 60% respectively, the difference would also be 10%. By multiplying the difference by $O(r)$, the difference in the first example remains at 10% while that in the second becomes to 500%. This adjustment results in highly shielded residues, as determined by RP-MS experiments, having a greater contribution to the final score of the complex.

The lower the score, the more closely the SAS shielding observed in the complex matches that measured in terms of oxidation. Thus a perfect match will result in a score of zero. Although not affecting the relative scores of complexes, the result obtained from Equation (2) is divided by the imperfect score $|S(r) - N_r|$, for $N_r$ and $S(r) = \max\{O(r) S(r)\}$ for $N_r$ and $N_c$ in order to generate percentile scores where 0% is a perfect match and 100% indicates a complex with the highest mismatch or imperfect score.

Using the default values (1.2 Å³ grid cubes, 1 Å thick surface layer and 15° rotation step), 6592 different complex conformations are generated during the geometric fitting routine. The 2000 structures displaying the greatest correlation according to Equation (1) are saved and this ensemble of conformations is scored and ranked according to several criteria. Most important of these is the shielding correlation as calculated in Equation (2). However, total SAS and energetic evaluation by Amber force fields (Wang et al., 2004) may also be used in scoring and ranking. Following scoring, complexes that have a RMSD of less than a cutoff value (5 Å default) are grouped into clusters in order to eliminate very similar results.

**IMPLEMENTATION**

**Ribonuclease S-protein–peptide complex**

The application of PROXIMO is first described for the ribonuclease S-protein–peptide complex. Ribonuclease A is a digestive enzyme secreted by the pancreas that, when cleaved by the protease subtilisin, produces the N-terminal S-peptide (residues 1–20) and S-protein moiety (residues 21–124). These can reassociate through hydrogen bonding and complexed forms. This is achievable if the proteolytic digest of the ribonuclease S complex was the first complex to be studied by RP-MS in which an electrical discharge source was employed as the means of producing hydroxyl radicals. The study demonstrated the selective shielding of reactive residues within S-peptide by S-protein when solutions containing both, together with a mixture of non-binding peptides, were subjected to radical-induced oxidation at a pH of 5.5 and 2. Upon formation of the ribonuclease-S complex at pH 5.5, a 77 and 55% reduction in the levels of oxidation were observed within S-peptide and residues 116–120 of S-protein, respectively.
Project details are entered into the GUI of the PROXIMO algorithm and the PDB files for the interacting molecules defined (Figure 2A). In this case, PDB files containing the coordinates for the two components of the complex, S-protein and S-peptide, are entered. The co-ordinates for S-protein were generated from those for the complex with co-ordinates for S-peptide removed. S-protein, being the larger of the two molecules, was selected as the static molecule while S-peptide was selected as the mobile molecule. PROXIMO then identifies the reactive residues and calculates the SAS of the reactive side chains. At this point, the user enters the shielding data obtained experimentally within the Oxidation Data window. Since the exact site(s) of oxidation in both S-peptide and S-protein residues 116–120 were not experimentally identified, common oxidation shielding values are entered for all reactive residues in each peptide. In this case, values for $O(r)$ are 54.97 for residues Phe-8, His-12 and Met-13 and 76.79 for residues Pro-117, His-119 and Phe-120 (Figure 2).

After data entry, the docking process was begun with settings for the geometric fitting routine assigned to default values. The 2000 conformations (numbered 0–1999) demonstrating the greatest surface shape complementarity, as evaluated by the geometric matching algorithm, were scored by PROXIMO based on the correlation of SAS shielding with the oxidation data. The resultant scores for the complexes ranged from 20.92 to 101.58. Of the 2000 hypothetical conformations, 97% received a score greater than 50, which in most cases corresponded to the ribonuclease–S protein–peptide interface being elsewhere than the reported binding domain. The top five scored complexes, with scores between 77.07 and 99.23 based upon geometric fitting alone, are shown in Figure 3 with S-protein illustrated in both a ribbon (3A) and surface (3B) representation.

The reactive residues in the 2000 saved conformations are identified and side chain SAS values are calculated. The structures of the complexes were then scored based upon a correlation between the oxidation shielding and SAS values according to Equation (2) and output into the Results window (Figure 2C). Following scoring, complexes that have a RMSD of less than a nominated cutoff value (5 Å default), are grouped into clusters in order to eliminate very similar conformations. All of the results are saved as PDB formatted files for later viewing. When any of the numbered conformations are selected in the first column of the left-hand pane, the SAS values for all shielded reactive residues appear in column three of the right-hand pane (Figure 2).

The top five complexes as scored by PROXIMO are shown in Figure 3C and D together with the shielding correlation score and RMSD between each structure and that determined by X-ray crystallography. All of the top five scored complexes (#157, 488, 113, 1175 and 217) had structures that deviated from the experimentally determined structure (crystal) with RMSD values between 0.45 and 1.26 Å$^2$ (Table 1). These values are all within the resolution obtained for the experimental structure (2.1 Å).

To demonstrate the ability of PROXIMO to successfully identify and score the experimentally determined structure, the PDB file representing coordinates for the experimental structure of the ribonuclease-S complex was combined with the 2000 complexes obtained from the geometric fitting routine. These were re-scored by the PROXIMO algorithm. The experimentally derived structure was ranked first in the output of results with a score of 18.75 and SAS shielding calculated at reactive residues is in accord with that input for oxidative shielding within the binding interface (Table 2). With the exception of His-119, the SAS shielding calculated for reactive residues in the native conformation of the ribonuclease-S complex displays a high correlation with the oxidation data obtained experimentally. The average SAS shielding across reactive residues in S-peptide is 72.76% that is within 5.25% of the oxidation shielding measured by RP-MS. The average SAS shielding in S-protein at residues Pro-117 and Phe-120 of 53.25% is within 3.13% of the oxidation shielding measured for the proteolytic peptide containing residues 116–120 of S-protein (Table 2).

The reduced correlation at histidine at position 119 is associated with its solvent accessibility within the complex. Since the sites of oxidation were not experimentally determined with the S-protein binding interface by RP-MS, a common value for oxidative shielding was entered for this residue and residues Pro-117 and Phe-120. If no oxidative shielding is entered for residue His-119, PROXIMO returns structures for the top five ranked complexes with much lower scores (<7) consistent with better fits between the oxidation and SAS shielding data (Figure 4). However, for the top three of these ranked complexes, S-peptide is incorrectly orientated in the structures numbered 46, 756 and 1224 and this leads to larger RMSD values from the experimental X-ray structure of between 3.6 and 4.4 Å$^2$. This result indicates the importance of establishing oxidation levels and shielding at each reactive residue, over a peptide domain as a whole, wherever possible.

**Calmodulin–melittin complex**

Calmodulin is a ubiquitous eukaryotic calcium-binding protein that interacts with and regulates a multitude of different protein targets (Crivici and Ikura, 1985). One of the many targets known to bind calmodulin with high affinity is the peptide melittin, a major component of bee venom. Upon binding with melittin, a dramatic conformational change in calmodulin occurs. The protein’s dumbbell structure consisting of two globular domains joined by a flexible $\alpha$-helix hinge region adopts a horseshoe-like structure in which
the N- and C-terminal domains, far removed in the protein, become proximal and engulf melittin within a hydrophobic channel (Babu et al., 1985). High-resolution experimental data are available for both calmodulin (Chattopadhyaya et al., 1992; Meador et al., 1992) and melittin (Terwilliger and Eisenberg, 1982) in their free form. Although there have been a large number of studies of the interaction between calmodulin and melittin, and even a report announcing the successful growth of crystals of the complex (Tanaka et al., 1985), no high-resolution data for the complex are available.

Analysis of the calmodulin–melittin complex has been studied by RP-MS within this laboratory. Solutions of calmodulin alone and in combination with melittin and several other non-binding peptides, at low and physiological pH, underwent radical-induced oxidation using an ESI discharge source. The partially oxidized products were subsequently digested with either trypsin or a mixture of trypsin and chymotrypsin. MS analysis of the products provided a measure of the extent of oxidation within calmodulin alone and in the presence of melittin. The extent of oxidation in the non-binding peptides was observed to remain constant whether in the presence or absence of calmodulin. However, there is a marked reduction of 46% (Table 3) in the oxidation of melittin while in the presence of calmodulin indicating complex formation and a corresponding shielding of specific reactive residues from the hydroxyl radical.

Table 1. Top five scored structures for the ribonuclease S complex using PROXIMO and the RMSD with experimentally determined (crystal) structure

<table>
<thead>
<tr>
<th>Structure</th>
<th>Score [from Equation (2)]</th>
<th>RMSD (Å²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>157</td>
<td>20.92</td>
<td>1.26</td>
</tr>
<tr>
<td>488</td>
<td>21.28</td>
<td>0.68</td>
</tr>
<tr>
<td>113</td>
<td>21.31</td>
<td>0.45</td>
</tr>
<tr>
<td>1175</td>
<td>26.37</td>
<td>0.99</td>
</tr>
<tr>
<td>217</td>
<td>28.01</td>
<td>0.91</td>
</tr>
</tbody>
</table>

*Score values range from 0% for a perfect match to 100% for the greatest mismatch. The lower the score the better the match between SAS shielding and oxidation shielding data.

Table 2. Correlation of SAS shielding of residue side chains at the binding interface in the ribonuclease S complex (native experimental) and the oxidative shielding measured by RP-MS

<table>
<thead>
<tr>
<th>Reactive residue</th>
<th>S(r) SAS shielding</th>
<th>O(r) oxidative shielding</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro-117</td>
<td>57.14</td>
<td>54.97</td>
</tr>
<tr>
<td>His-119</td>
<td>0</td>
<td>54.97</td>
</tr>
<tr>
<td>Phe-120</td>
<td>49.35</td>
<td>54.97</td>
</tr>
<tr>
<td>S-peptide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe-8</td>
<td>80.95</td>
<td>76.79</td>
</tr>
<tr>
<td>His-12</td>
<td>60.83</td>
<td>76.79</td>
</tr>
<tr>
<td>Met-13</td>
<td>76.50</td>
<td>76.79</td>
</tr>
</tbody>
</table>

Oxidation levels were also measured within the proteolytic peptides of calmodulin. Marked reductions were observed across residues towards the N- and C-termini consistent with the low-resolution model for the complex. A dramatic decrease in oxidation of 61.5 and 68.4% was observed for peptides containing residues 14–21 and 95–106, respectively, that was localized to residues Phe-16 and 19, and Tyr-99. An increase in oxidation of 7.6% was measured at Phe-92, a phenomenon associated with its greater accessibility to solvent and the hydroxyl radical upon the conformational change of calmodulin.

Measured oxidation protection at these residues together with shielding values of 0 for reactive residues Met-36, Phe-89, Tyr-138, Phe-141, Met-144 and 145 were entered into the PROXIMO algorithm together with co-ordinates for the structures of calmodulin and melittin. The 2000 conformations from the geometric fitting routine alone contained conformations that were far removed from those predicted for the complex and were ranked by PROXIMO using the oxidation data and assigned scores ranging from 34.76 to 63.44. Since these conformations were produced through rigid body docking of the structures of calmodulin and melittin, based on co-ordinates derived from their respective PDB files (1CLL and 2MLT, respectively), not one of the top ten structures had melittin centred about the N- and C-termini as predicted by the horseshoe model for the complex (Scaloni et al., 1998). Implementation of the molecular dynamics routine on the top five and other sets of scored conformations also failed to predict a structure for the complex that was closer to the model. Structural clashes, resulting from an overlap of atoms in common space, remained unresolved even after energy minimization. Clearly,
allow for some penetration of melittin into the hydrophobic channel. The combined shieldings are in accord with oxidative shielding of
62.12 (Table 3). Phenylalanine at position 19 is shielded between 55.45 oxidation shielding measured in RP-MS experiments of 45.87 s
27.34 and 40.07 tryptophan residue at position 19 in melittin is shielded by between 32.95 and 36.35. Four of the five structures have melittin at
least partially buried within the hydrophobic channel (Figure 5). The large conformational change that calmodulin undergoes upon its
binding to melittin makes predicting a near native structure of the calmodulin–melittin complex much more difficult than in the case
of the ribonuclease-S complex even when the built-in energy minimization and molecular dynamics routine within PROXIMO is utilized (Figure 2).

A second docking strategy was employed in which the structure for the complex of calmodulin bound to the MCLK (Myosin light chain kinase) derived peptide (PDB file 1CDL) was used to obtain co-ordinates for calmodulin alone in its globular folded state. These co-ordinates and those for melittin were entered into PROXIMO together with the oxidation shielding data using the same docking parameters as above, except the surface layer thickness s was raised to 2Å to allow for some penetration of melittin into the hydrophobic channel.

The algorithm outputs the top five ranked complexes with scores between 32.95 and 36.35. Four of the five structures have melittin at least partially buried within the hydrophobic channel (Figure 5). The tryptophan residue at position 19 in melittin is shielded by between 27.34 and 40.07 Å² across the five structures in accord with the oxidation shielding measured in RP-MS experiments of 45.87 (Table 3). Phenylalanine at position 19 is shielded between 55.45 and 62.12 Å² while Phe-16 is shielded between 2.94 and 11.76 Å². The combined shieldings are in accord with oxidative shielding of

61.5% derived from RP-MS data. This results in the respective exposure of the side chain of residue Phe-16 with solvent accessibilities of 2.94–11.76 Å². Thus a correlation between these four structures and RP-MS data obtained is found where oxidation largely occurs within the peptide comprising residues 14–21 or only at Phe-19.

These same four structures, however, are not reconciled simultaneously with the reduction in oxidation observed at residue Tyr-99 in peptide 95–106. A reduction in oxidation of 68.4% is measured in RP-MS experiments, in accord with the interaction of the melittin with the surface of calmodulin elsewhere to the channel that accommodates the MLCK peptide. Only one of the top five scored structures, namely 1688 (Figure 5), shields the side chain of this tyrosine residue in accord with the RP-MS data. A SAS shielding S(r) value of 72.55 is recorded for this residue in structure number 1688 but is between 0 and 2.88 in all other structures (Table 3).

The results illustrate that the structure for the calmodulin–melittin complex has not been correctly modelled by the algorithm since oxidation shielding measured at residues Phe-16, Phe-19 and Tyr-99 was not reconciled simultaneously in any of the top five structures. This indicates that the structure of calmodulin derived from the calmodulin–MLCK peptide complex is significantly different to that of the calmodulin–melittin complex. This result is interesting in its own right in that RP-MS data together with the PROXIMO algorithm can dismiss structures for protein complexes that would be predicted from docking by a geometric fitting routine alone. A greater conformational change toward the C-terminus of calmodulin in the structures 132, 437, 1352 and 1905 is required to shield the side chain of tyrosine at position 99 while still accommodating melittin in a hydrophobic channel to shield the side chain of Trp-19. Despite repeated experiments, this conformational change was too great to be modelled with molecular dynamics simulations using PROXIMO though does not invalidate the use of the algorithm for docking protein complexes in which less pronounced conformational changes occur.

CONCLUSIONS

The design and implementation of the PROXIMO algorithm has been employed to model the structures for the ribonuclease S and calmodulin—melittin complexes with the use of oxidation shielding data obtained from radical probe mass spectrometry (RP-MS) experiments. The structure of the ribonuclease S complex is in close agreement with its experimental high-resolution structure illustrating that the algorithm, in conjunction with RP-MS data, will

<table>
<thead>
<tr>
<th>Structure</th>
<th>S(r) SAS shielding of reactive residue side chain</th>
<th>Ox(r) oxidation shielding</th>
</tr>
</thead>
<tbody>
<tr>
<td>1352</td>
<td>2.88, 2.94, 61.21, 32.31, 45.87</td>
<td>68.4</td>
</tr>
<tr>
<td>1905</td>
<td>1.90, 11.76, 55.45, 27.34, 40.07</td>
<td>61.5</td>
</tr>
<tr>
<td>1688</td>
<td>72.55, 4.41, 3.33, 31.62, 45.87</td>
<td>61.5, 45.87</td>
</tr>
<tr>
<td>437</td>
<td>1.27, 5.88, 62.12, 40.07</td>
<td>0, 2.88 in all other structures (Table 3).</td>
</tr>
<tr>
<td>132</td>
<td>0, 7.35, 59.69, 35.70, 45.87</td>
<td>0, 2.88 in all other structures (Table 3).</td>
</tr>
</tbody>
</table>

Table 3. Solvent accessibility surface shielding of reactive residue side chains versus oxidative shielding from RP-MS data in the top five scored structures for the calmodulin–melittin complex.

Figure 4. Ribbon (top) and surface (bottom) representations of the top five scored structures for the Ribonuclease S complex where the shielding at residue His-119 was set to zero. One terminus of S-peptide is labelled to show its orientation.

CONCLUSIONS

The design and implementation of the PROXIMO algorithm has been employed to model the structures for the ribonuclease S and calmodulin—melittin complexes with the use of oxidation shielding data obtained from radical probe mass spectrometry (RP-MS) experiments. The structure of the ribonuclease S complex is in close agreement with its experimental high-resolution structure illustrating that the algorithm, in conjunction with RP-MS data, will
predict structure for protein complexes. Large protein complexes that are difficult to study experimentally are particularly suited to this approach since the size of each binding partner imposes more constraints during the docking process where there is no significant conformational change.

In the case of a binding partner undergoing a more significant conformational change during complexation, the algorithm is more challenged but nonetheless may yield useful results where the oxidation data in RP-MS experiments is obtained for significant numbers of reactive residues.

The algorithm, which incorporates a simple to use GUI, is predicted to advance the wider application of RP-MS experiments and assist with data interpretation. It can also be utilized with other quantitative shielding data types to assist with data-driven protein docking more generally.

ACKNOWLEDGEMENTS

S.K.G. thanks the Biochemical ALgorithm Library (BALL) development team, and especially Oliver Kohlbacher and Andreas Hildebrandt, who provided a pre-release version of the library and who answered questions concerning implementation of the software library.

Conflict of Interest: none declared.

REFERENCES