J Biomol NMR (2009) 43:11–19
DOI 10.1007/s10858-008-9286-7

ARTICLE

Use of quantitative ^{1}H NMR chemical shift changes for ligand docking into barnase

Marina Cioffi · Christopher A. Hunter · Martin J. Packer · Maya J. Pandya · Mike P. Williamson

Received: 28 July 2008 / Accepted: 9 October 2008 / Published online: 1 November 2008

Abstract ^{1}H NMR complexation-induced changes in chemical shift (CIS) of HN protons have been used to characterize the complexes of barnase with the deoxyoligonucleotides d(GC) and d(CGAC). Quantitative shift changes are used not only to locate the most probable binding site (using ring-current shifts), but also to determine the orientation of the ligand within the binding site, based on a more complete shift calculation including bond magnetic anisotropies and electric field effects. For both ligands, the guanine is in the same binding site cleft, in the same position as identified in the crystal structure of the d(CGAC) complex. By contrast, a previous X-ray crystal structure of the d(GC) complex showed the ligand in the mouth of the active site, rather than at the guanyl-specific site, implying that the location may be an artifact of the crystallisation process.

Keywords Barnase · Complexation induced shift · Crystal packing · Docking · Guanine

Abbreviations

CIS	Complexation-induced shift
$\Delta\delta_{exp}$	Limiting chemical shift change on addition of ligand

M. Cioffi · C. A. Hunter
Department of Chemistry, University of Sheffield, Dainton Building, Brook Hill, Sheffield S3 7HF, UK

M. J. Packer
AstraZeneca, Alderley Park, Cheshire SK10 4TG, UK

M. J. Pandya · M. P. Williamson (✉)
Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, UK
e-mail: m.williamson@sheffield.ac.uk

Introduction

Chemical shift mapping is a technique very widely used to determine the binding site of a ligand on a protein, in which chemical shift changes in the protein (usually of the amide protons and nitrogens) on addition of the ligand are measured, and the largest changes are used as an indicator of the likely binding site. This technique is simple and flexible, because the magnitude of shift taken to be 'significant' is user-defined, thus allowing one to choose a value that gives a structurally meaningful group of residues. It is also rapid, making it competitive with crystallography as a method for identifying ligand binding sites. Residues or atoms identified as 'significantly shifted' can then be used in docking procedures, most commonly in manual docking or alternatively using an algorithm such as the widely used HADDOCK, in which residues with shift changes are used as the basis for a set of ambiguous intermolecular distance restraints (Dominguez et al. 2003; Schieborr et al. 2005). Such methods have been usefully described as 'information-driven docking' (van Dijk et al. 2005).

It is however, possible to use shifts in a more quantitative way. Morelli et al. (2000, 2001) incorporated a filter based on a number of parameters including chemical shift changes in the program BiGGER (Palma et al. 2000). Other docking programs have used shifts in a similar way as scoring functions to rank solutions from conformational searches (Dobrodumov and Gronenborn 2003; Kohlbacher et al. 2001; Stark and Powers 2008). A very interesting methodology was described by McCoy and Wyss (McCoy and Wyss 2002). They noted that most ligands that are useful as pharmaceutical agents contain aromatic rings as key elements of the binding interface; and furthermore that the limiting chemical shift change (Complexation Induced Shift or CIS) caused by an aromatic ring takes the form

$$\Delta\delta_{exp} = (f[\theta])/r^3 \quad (1)$$

where θ is the angle between the affected proton and the centre of the aromatic ring, and r is the distance. This implies that the *largest* possible distance from the centre of the ring to the affected proton is directly proportional to $\Delta\delta_{exp}{}^{-1/3}$, and therefore that any measured ring-current shift change can be used to locate the ligand within a sphere of radius proportional to $\Delta\delta_{exp}^{-1/3}$. They therefore suggested drawing such spheres around each perturbed proton, and defining the location of maximum overlap (maximum probability density) as the most likely ligand binding site. They wrote a program based on this idea, called *Jsurf*. We have adopted this method here, and incorporated it into a three-stage procedure.

The first stage of our procedure uses *Jsurf* to identify the most likely position for the 'centre' of the ligand. We then use the docking program *GOLD* (Jones et al. 1997) to generate a large set of possible sterically allowed ligand orientations, based on the ligand position identified in *Jsurf*. Finally, we optimise the ligand orientations by making a more complete calculation of the chemical shift effect of the ligand, and use this to identify the ligand positions that produce the best fit to the experimentally measured shifts. The method has been described elsewhere (Cioffi et al. 2008a): here we describe an application of the method to a real problem, namely the docking of oligonucleotides into barnase.

Barnase has been extensively used as a model protein (Meiering et al. 1993), and has 110 amino acids (MW = 12.4 kDa) with no disulfide bridges. It is a small endoribonuclease, and catalyses the cleavage of single stranded RNA, using His102 and Glu73 as general acid and base, respectively. The inactive His102Ala mutant has been widely used in place of the wild type, as it is simpler to express (Mossakowska et al. 1989).

The binding of barnase with several substrate analogues such as mono-, di- and tetra-deoxynucleotides has been investigated by NMR spectroscopy and X-ray crystallography. These studies show that barnase is guanyl-specific for short oligonucleotides, but less specific for oligonucleotides more than 4–5 bases in length. This decrease in specificity has been attributed to secondary binding sites (Day et al. 1992; Mossakowska et al. 1989). In particular, there are crystal structures of both the 3′-GMP complex (Baudet and Janin 1991) and the complex with d(CGAC) (Buckle and Fersht 1994). The guanines are in similar positions in both, within the active site. By contrast, in the crystal structure of the d(GC) complex, the nucleotides occupy a quite different position in the mouth of the active site, in a very similar position to that occupied by the AC dinucleotide of d(CGAC). It has been suggested that this alternative location may be due to crystal packing (Baudet and Janin 1991), but there have been no NMR studies carried out to investigate the position in solution. In this paper, we describe solution NMR binding studies and chemical shift-based structure determinations on the complexes of barnase with d(GC) and d(CGAC), demonstrating that both bind in the same orientations. The method described here is robust and flexible (although limited to aromatic ligands), and should have wide applicability.

Materials and methods

Purification of barnase

Escherichia coli strain M15 [pREP4] cells (Qiagen) transformed with the plasmid pQE-60 (Qiagen) expressing barnase H102A were grown at 37°C in M9 minimal medium containing ampicillin (100 μg/ml), kanamycin (100 μg/ml), ^{15}N-ammonium sulphate and ^{13}C-glucose. Expression was induced by the addition of 1 mM isopropyl-β-D-thiogalacto-pyranoside (Melford), and the culture was incubated for 18 h at 25°C and harvested by centrifugation (5,000*g* for 20 min at 4°C). Cells were disrupted in 25 mM Tris–HCl, pH 7.4, containing 5 mM $MgCl_2$ and ethylenediaminetetraacetic acid-free complete protease inhibitor mix (Roche) and the extract was centrifuged (40,000*g* for 40 min at 4°C) after addition of 25 U ml^{-1} Benzonase nuclease (Sigma). The supernatant was dialysed (Spectra/Por membrane, molecular weight cut-off 3,500) against 25 mM Tris–HCl, pH 7.4, containing 0.02% Na azide at 4°C, and applied to a column (1.5 × 21 cm) of Q Sepharose Fast Flow (G E Healthcare) equilibrated with the same buffer. After washing the column with buffer, nuclease-free barnase was collected and dialysed against 50 mM Na acetate buffer, pH 5, containing 0.02% Na azide at 4°C. The protein was then applied to a column (1.5 × 18 cm) of SP Sepharose Fast Flow (G E Healthcare) equilibrated with the same buffer. After washing the column with starting buffer, a linear gradient of 0.1 M NaCl over 500 ml was applied.

Barnase purity was confirmed by tris-tricine sodium dodecyl sulphate polyacrylamide gel electrophoresis in the presence of reducing agent, 2-mercaptoethanol (Schägger and von Jägow 1991). Protein was dialysed against 30 mM Na phosphate buffer, pH 6.7, containing 0.02% Na azide at 4°C, then concentrated with an Amicon stirred cell (Ultracel PLBC 3000 ultrafiltration membrane, Millipore). Protein stocks were stored at −80°C and concentrations were estimated by absorption at 280 nm. The identity and integrity of the purified barnase was confirmed by matrix assisted laser desorption/ionization time-of-flight mass spectrometry and amino-terminal sequencing.

The protein samples used for titration experiments were prepared by dissolving the lyophilized protein (^{13}C–^{15}N barnase H102A) in 90% H_2O and 10% D_2O (30 mM NaPhosphate, NaN_3 0.02%) at pH 7.

NMR titration experiments

^{1}H-NMR and ^{1}H–^{15}N HSQC spectra were acquired at 298 K for the protein sample alone and after each addition of the ligand, on a 600 MHz Bruker spectrometer for the titration with d(GC) and on a 500 MHz Bruker spectrometer for the titration with d(CGAC), and processed using FELIX (Accelrys Inc, San Diego CA). The ^{1}H–^{15}N assignment for the free barnase was carried out by reference to literature assignments (Korzhnev et al. 2001). This assignment of the free protein was used as a starting point to obtain the full assignment of the spectrum recorded after each addition. For barnase-d(GC), 7 aliquots (10, 10, 10, 10, 20, 40, 80 µl) of ligand stock d(GC) (2.5 mM in D_2O/buffer) were added to the protein sample (200 µM in 500 µl) at pH 7 to reach a final ratio of barnase:d(GC) of 1:4.5. For barnase-d(CGAC), three aliquots (each of 10 µl) of ligand d(CGAC) stock (5 mM in D_2O/buffer) were added to the protein sample (200 µM in 500 µl) at pH 7 to reach a final ratio of barnase:d(CGAC) of 1:1.5. The largest observed ^{1}H shift changes were fit to a 1:1 binding isotherm using a purpose-written program in Microsoft Excel to obtain the association constant and the limiting complexation-induced change in chemical shift. This allowed calculation of a speciation profile to estimate the fraction bound at each point in the titration, and hence limiting complexation-induced chemical shift changes for all of the signals. For each signal in each spectrum in the titration experiment, the limiting experimental shift change $\Delta\delta_{exp}$ was calculated using Eq. 2. Where individual measurements deviated from the average value for a particular signal by more than two standard deviations, they were considered outliers and were eliminated. The remaining values of $\Delta\delta_{exp}$ were then used to calculate an average value of $\Delta\delta_{exp}$ at saturation.

$$\Delta\delta_{exp} = \frac{\Delta\delta_{obs}}{f_{bound}} \quad (2)$$

where $\Delta\delta_{obs}$ is the chemical shift observed in a particular experiment and f_{bound} is the fraction bound determined using the signals that showed large changes in chemical shift and could be fit reliably to a 1:1 binding isotherm.

Determination of the structure of the complexes

The coordinates of the protein frames used in the calculations were obtained from the protein data bank (PDB entry code 1BNR for the NMR solution structure of the unbound protein, 1RNB for the X-ray crystal structure of the protein bound to d(GC) and 1BRN for the X-ray crystal structure of the protein bound to d(CGAC)). The structure of the ligand was created with XED 6.1.0 (Vinter 1996) using standard bond lengths and angles and energy minimised. Our computational approach consists of a set of Perl scripts and C++ programs that implement the three main software packages used in the protocol and analyse the results. The three programs used are *Jsurf* to define the receptor binding site using the backbone amide CIS values, *GOLD* to generate a set of ligand conformations and orientations ('poses') for introduction into the receptor binding site, and *Shifty* to optimise each pose based on comparison of the experimental and calculated CIS values for the amide backbone protons.

Jsurf

In the first stage of the protocol, the location of the binding site was obtained using the program *Jsurf* (McCoy and Wyss 2002). Experimental CIS values of the backbone amide protons were first mapped onto the protein van der Waals surface, to check for the existence of a single defined binding site. The protein frame was coloured depending on the intensity of the experimental CIS values. The largest CIS values were found to be clustered in the natural cleft of the protein, and no significant perturbations were identified elsewhere on the protein surface. A ligand *j*-surface was therefore created. Each significantly shifted proton is used to define a sphere of radius proportional to $\Delta\delta_{exp}{}^{-1/3}$. The sphere is composed of a number of randomly placed dots, in which each dot represents a possible location for the centre of the ligand. Because there are the same number of dots in each sphere, larger chemical shift changes generate smaller spheres containing a greater density of dots. The regions of highest 'dot density' (i.e. regions of large shift change and in which spheres from several protons overlap) are identified using a user-defined threshold, calculated as *n* standard deviations above the mean. In the calculations carried out here, *n* was varied between 1 and 4, but the best value was found to be around 2. Values that were too small produced two separate small surfaces. This may be due to the presence of two different binding sites or more likely to the fact that the ligand has several aromatic groups. Regions of density greater than the threshold are connected to form a *j*-surface, containing the most likely positions of the ligand centre. All points on the *j*-surface less than 2.5 Å from the protein backbone were removed, and the remaining points were averaged to give the coordinates of a single point: the centre of the binding site.

GOLD

In the second stage, the generation of an ensemble of poses located in the binding site was carried out using *GOLD* v.2.2. *GOLD* is a search and optimisation docking program based on a genetic algorithm, which generates a set of good fits of a ligand into its binding site, measuring goodness of fit using user-defined scoring functions (Jones et al. 1997). The ligand is allowed free rotation around defined bonds, while the protein may have flexibility in sidechains close to the ligand binding site. The centre of the ligand was constrained at the centre of the binding site determined using *Jsurf*. The *GOLD* scoring function was set so as to produce rapidly a set of poses based on shape complementarity only, by setting the contribution of the hydrogen bond energy term very low, switching off the molecular interaction terms, and considering only van der Waals and intramolecular strain energies. *GOLD* was used to generate structures using 10 runs with a population size of 100 for 10,000 generations. A diverse set of poses was generated by switching off the 'early termination' option (commonly used to terminate runs as soon as a specified number of runs have given essentially the same answer).

Shifty

The final stage of the structure determination used the diverse set of structures produced by *GOLD* as starting points for a conformational search, again carried out using a genetic algorithm, which uses chemical shift as its sole fitness criterion. The fitness of a particular structure is defined by the normalized root mean square difference between the experimental and calculated CIS values (weighted by a global scaling factor designed to reflect the fact that the ligand is not necessarily 100% bound to the protein). The effect of the ligand on amide proton chemical shifts in the protein is calculated using optimised parameters based on ring currents, bond magnetic anisotropy and electric field effects. This calculation was carried out using the program *Shifty* (Cioffi et al. 2008a; Cioffi et al. 2008b; Hunter and Packer 1999). The conformational search in *Shifty* was divided into two steps, each with population sizes of 50 run for 50 generations, in which the protein was kept rigid but the ligand was allowed to undergo torsional rotations. The size of the search space was set to ensure that all possible complex conformations could be sampled. In the first step, the intermolecular distance limit was set to 1.5 Å, and the range of allowed rotations of one molecule relative to the other was set to ±10°. Intramolecular torsions were allowed to change within the full range of ±180°. In the second step, these parameters were reduced to 1 Å, 5 and 90°, respectively. To reduce the conformational space, a steric clash penalty was added for distances of less than 2 Å for intermolecular clashes and distances less than 1 Å for intramolecular clashes for non-hydrogen atoms.

Results

NMR titration experiments were carried out to study the binding of the protein barnase with the nucleotides d(GC) and d(CGAC) (Fig. 1). The protein sample used for the titrations was a mutant of the wild-type barnase (H102A) which is inactive.

Titrations

Titration of d(GC) into barnase was monitored using $^{1}H–^{15}N$ HSQC spectra. Some signals could not be assigned due to overlap (Leu14, Ser28, Ala74, Tyr78, Phe82, Asn84 and Ile109) or broadening (Gly52, Ser57 and Asn58). The ^{1}H δ_{obs} for the remaining residues that were affected most during the titration (Ala102, Glu62, Tyr103, Leu42, Asp54, Lys62 and Glu60) were fit to a 1:1 binding isotherm (Fig. 2). The fitting procedure yielded a dissociation constant of 0.59 mM, which indicated that 50% saturation was achieved in the titration. This allowed calculation of the predicted limiting changes in chemical shift for formation of the 1:1 complex for all of the signals.

The titration results for the d(CGAC) complex were analysed in the same way as for the d(GC) complex

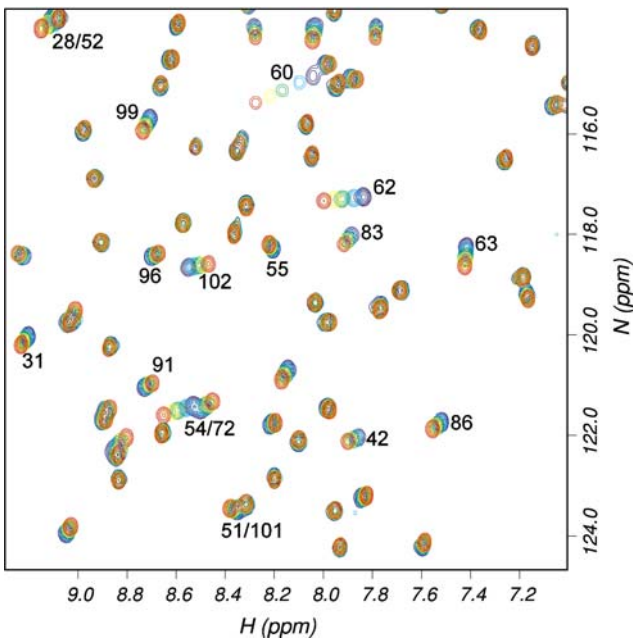

Fig. 1 Overlay of $^{1}H–^{15}N$ HSQC spectra of barnase H102A complexed with d(GC) at different concentrations. Cross peaks that shift most upon binding are labelled by residue number

(Fig. 3), yielding a dissociation constant of 49 μM, implying that 86% saturation was achieved in the titration. We note that these dissociation constants are similar to but slightly weaker than the K_m values obtained for wild-type barnase against the corresponding ribooligonucleotides, as expected (Day et al. 1992).

Structure determination

The experimental CIS values were used to calculate the structures of the complexes using a three-stage procedure described in detail elsewhere (Cioffi et al. 2008a; Cioffi et al. 2008b). Briefly, the first stage was to use the program *Jsurf* to calculate the most probable location of the centre of the ligand, assuming that all chemical shift changes observed on the amide protons of the protein originate from ring current shifts produced by the ligand. The second stage is to use the program *GOLD* to generate a large number (typically 100) of sterically allowed orientations ('poses') of the ligand, centred at the location indicated by *Jsurf*. The third stage is to use an optimisation procedure based on a reasonably full chemical shift calculation to determine the best location of the ligand, starting from each of the poses calculated by *GOLD*. This method was applied to both the d(GC) and d(CGAC) complexes.

Barnase-d(GC)

The X-ray crystal structure of the barnase-d(GC) complex (PDB code 1RNB) was used as the protein input frame. The ligand structure (in mol2 format) was built using molecular mechanics and energy minimized (Vinter 1996). When the protein surface was coloured according to the magnitude of the CIS values, it was readily observable that the largest changes were located in a small area of the protein, whereas the rest of the residues did not show any significant changes. This demonstrates, as reported in the literature, that the CIS observed are mainly due to the binding of the ligand, and that the protein does not undergo any large conformational rearrangement upon binding. Therefore in the first step of the procedure, the experimental CIS values were mapped onto the van der Waals surface of the protein structure using *Jsurf* (McCoy and Wyss 2002) (Fig. 4a). Using dot-density representations, spheres were constructed centred on each perturbed proton to create a *j*-surface representation of the binding site, in which each dot represents a possible location for the centre of the ligand (Fig. 4b). The most probable location of the ligand was identified by calculating the regions of higher dot density, removing any regions inside the protein surface, and averaging the remaining points to give a single point identifying the centre of the binding site (Fig. 4c). This provides a good starting point for location of the centre of the ligand, obviating the need for searching a large amount of redundant conformational space remote from the binding site.

This location was used to generate a set of starting conformations of the ligand in the protein binding site using *GOLD*, which generated a set of 100 poses based on shape complementarity only. Each of the poses generated by *GOLD* was then used as an independent starting point

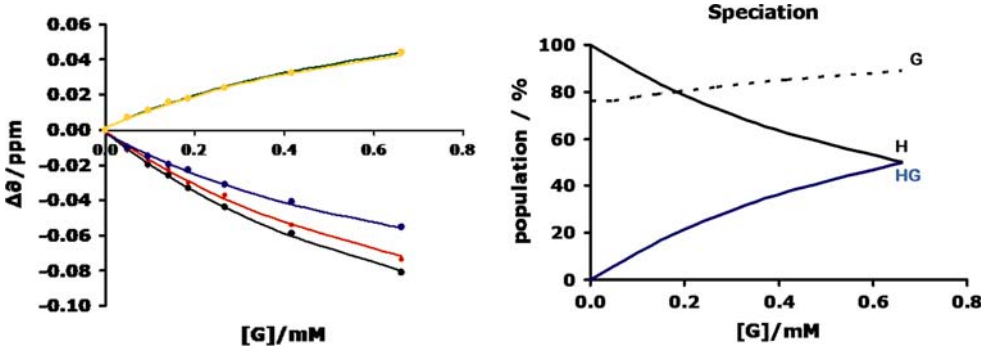

Fig. 2 ^{1}H NMR titration data for addition of d(GC) to barnase. The best fit to a 1:1 binding isotherm is shown (left, for five selected residues) along with the corresponding speciation profile (right). H is the population of free barnase, G the population of free ligand and HG the population of complex

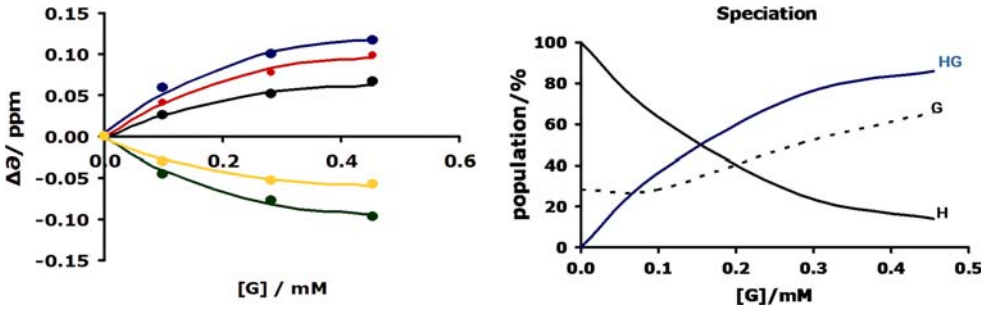

Fig. 3 ^{1}H NMR titration data for addition of d(CGAC) to barnase. The best fit to a 1:1 binding isotherm is shown (left, for five selected residues) along with the corresponding speciation profile (right). H is the population of free barnase, G the population of free ligand and HG the population of complex

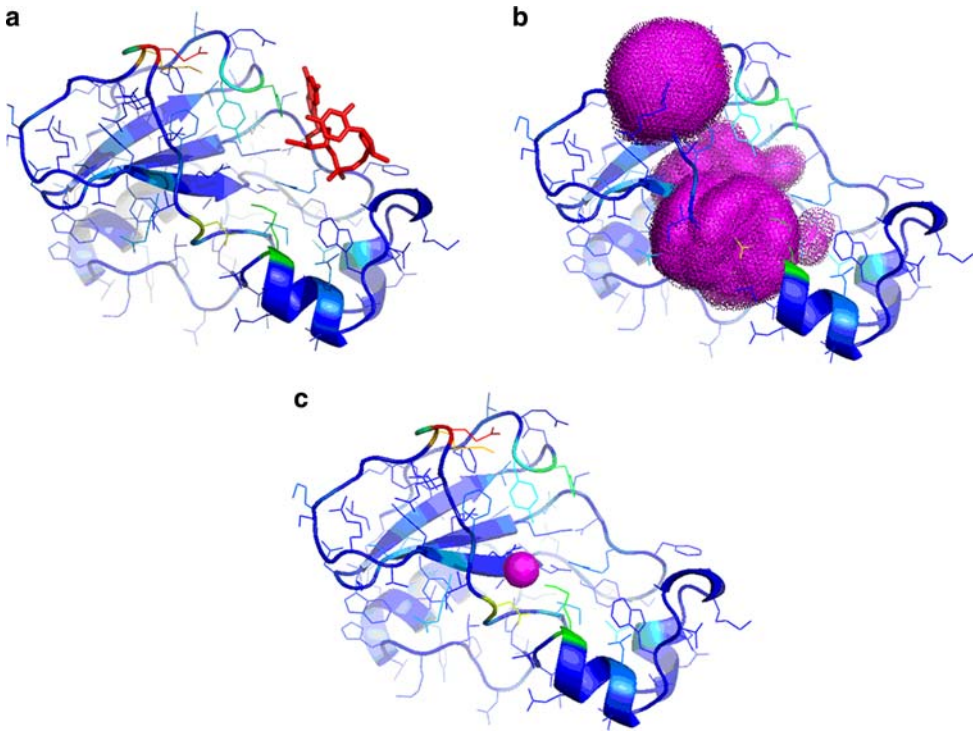

Fig. 4 **a** Experimental CIS mapped onto the X-ray crystal structure of the barnase-d(GC) complex. The ligand is shown in red sticks, and the protein is coloured according to the magnitude of the CIS values. The largest absolute changes are represented in red, moderate changes in yellow and green and smallest changes in cyan. **b** *j*-surface (magenta) representing potential locations for the centre of the ligand. **c** Centre of the *j*-surface

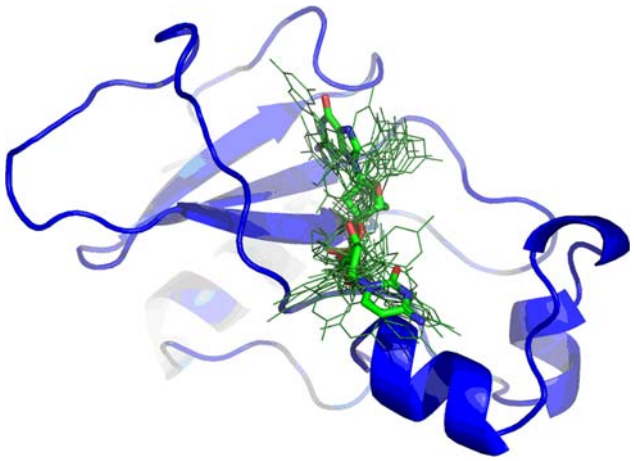

Fig. 5 Overlay of the highest fitness structures of the barnase-d(GC) complex after optimisation using the CIS values. The structure in bold is a typical structure, close to the average of the high fitness structures

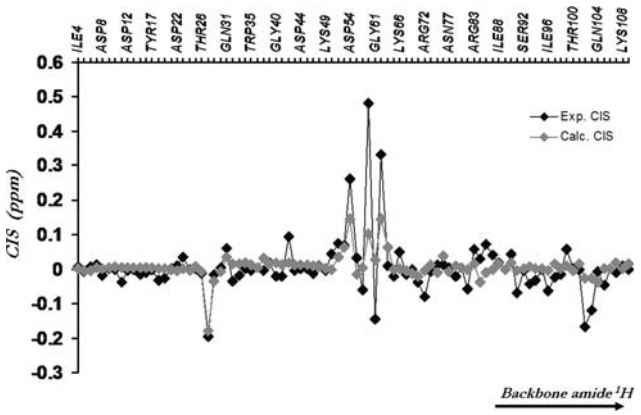

Fig. 6 Comparison between calculated and experimental CIS values of the protein backbone amide protons for the barnase-d(GC) complex

and optimised using the experimental CIS data using the program *Shifty*.

As expected, *GOLD* generated a wide range of ligand orientations. However, the optimisation process within *Shifty* was able to converge to a clearly defined structure (Fig. 5), with good agreement between the experimental and calculated backbone amide CIS values (Fig. 6). This location of the ligand is a poor match to the corresponding crystal structure, being 6.74 Å RMS from the location seen in the crystal structure (Fig. 7), but the guanine of the ligand is in an orientation that is similar to the binding mode observed by X-ray crystallography for 3′-GMP and d(CGAC) (RMSD for guanine = 2.94 Å, Fig. 7). The van der Waals surface of the ligand fits well into the protein surface, especially around the aromatic rings (Fig. 7b), as expected since this is the main criterion of the *GOLD* docking procedure. Similar results were obtained when using an unbound NMR solution structure of the protein (Bycroft et al. 1991) as the protein input frame (RMSD for guanine = 3.52 Å).

Some peak assignments were ambiguous and were not used in our calculations. However, some of the unassigned signals moved significantly during the titration. Therefore, after calculating the optimised structure we went back to

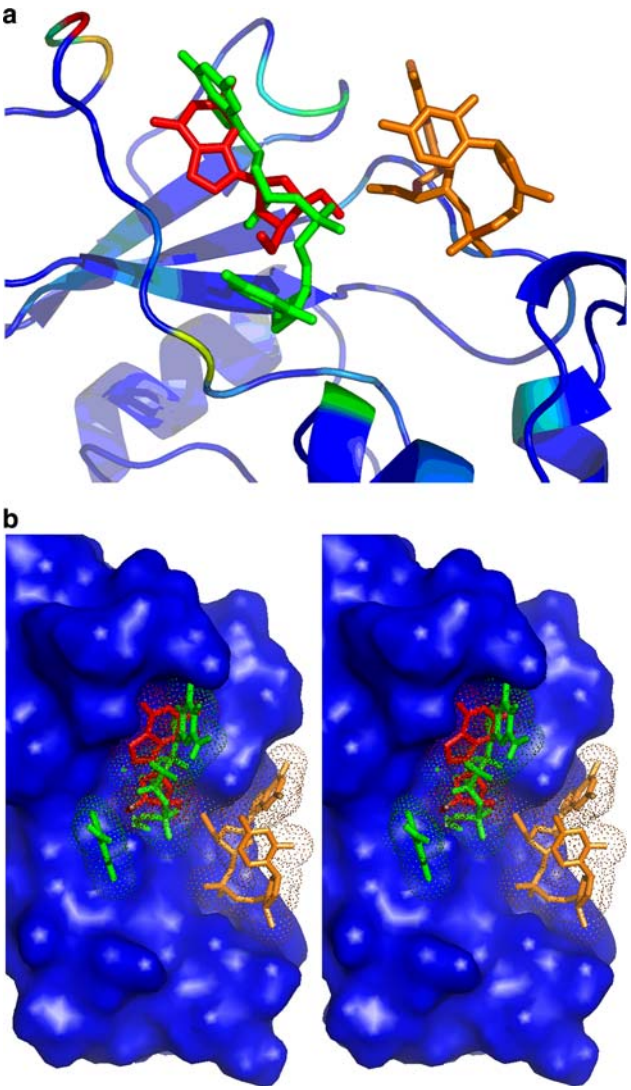

Fig. 7 **a** Superposition of the optimised CIS structure of the barnase-d(GC) complex (green) with the ligand in the crystal structure of the barnase-d(GC) complex (orange) and with the crystal structure of the guanine from the Barnase-d(CGAC) complex (red). **b** Same superposition viewed from a slightly different angle as a stereo view, with the protein van der Waals surface shown in blue and the ligand surfaces as dots

the spectra and compared experimental and calculated shifts. In nearly all cases the experimental changes in shift are at locations predicted by *Shifty* to move, thereby permitting an assignment of ambiguous shifts and potentially providing additional data for the fitting.

Barnase-d(CGAC)

The solution binding mode for the barnase-d(CGAC) complex was determined in the same way. Two different protein frames were used as the input to the structure determination procedure: the NMR solution structure of unbound barnase (PDB code 1BNR) and the X-ray crystal structure of barnase bound to the dinucleotide d(GC) (PDB code 1RNB). A *j*-surface was created (Fig. 8) and fed into *GOLD* to generate a set of poses, which were optimized using *Shifty*. There is good agreement between experimental and calculated backbone amide CIS values (Fig. 9), and the guanine nucleotide overlaid well with the position of the guanine in the d(CGAC) crystal structure (Fig. 10): the RMSD was 2.96 Å in the structure calculated using the free solution structure and 3.80 Å using the crystal structure of the d(GC) complex.

Discussion

Most applications of CIS to docking have used the shift changes in a purely qualitative way: shift changes are grouped into 'significant' and 'insignificant' categories and the ligand is docked to be adjacent to the 'significant' shifts, either manually or using some kind of ambiguous restraints (Dominguez et al. 2003). That is, shifts are used to guide the docking. This approach is simple to understand and simple to do, but is inherently unsatisfying because it throws away a large amount of potentially useful quantitative data. More quantitative approaches have generally used the numerical values of shift changes as a filter. By contrast, in our method the shift changes are used as restraints for optimisation of the family of poses produced by the combined use of *Jsurf* and *GOLD*. The results presented here demonstrate that this is a method capable of producing good geometries of the bound complex. In the method described here, *GOLD* was used to generate poses based purely on van der Waals complementarity, and there were no hydrogen bonds or electrostatic terms. This means that the structures obtained from *GOLD* are not biased by the empirically parameterised terms used to describe intermolecular interactions and are based on shape complementarity only. It is noteworthy that equally good results were obtained starting from either the *apo*- or *holo*-protein structure. In most applications only the *apo*-structure will be available. Provided that there are no gross changes in protein structure on ligand binding, this appears to be a good starting point (Cioffi et al. 2008b).

The initial search is very simple, in that we average the points within the *j*-surface to obtain the most likely centre of the ligand. Clearly, if the ligand is large or contains more than one aromatic ring, the 'average' position obtained may be inaccurate. A further problem is illustrated by Fig. 4: on binding of barnase to d(GC) there are some large chemical shift changes near the top of the protein in the orientation shown in Fig. 4, and it is clear from the final structure of the complex that these are not caused by ring current shifts from the ligand at all, but are in fact due to a conformational change of a protein loop. Thus in general we may expect the initial ligand location

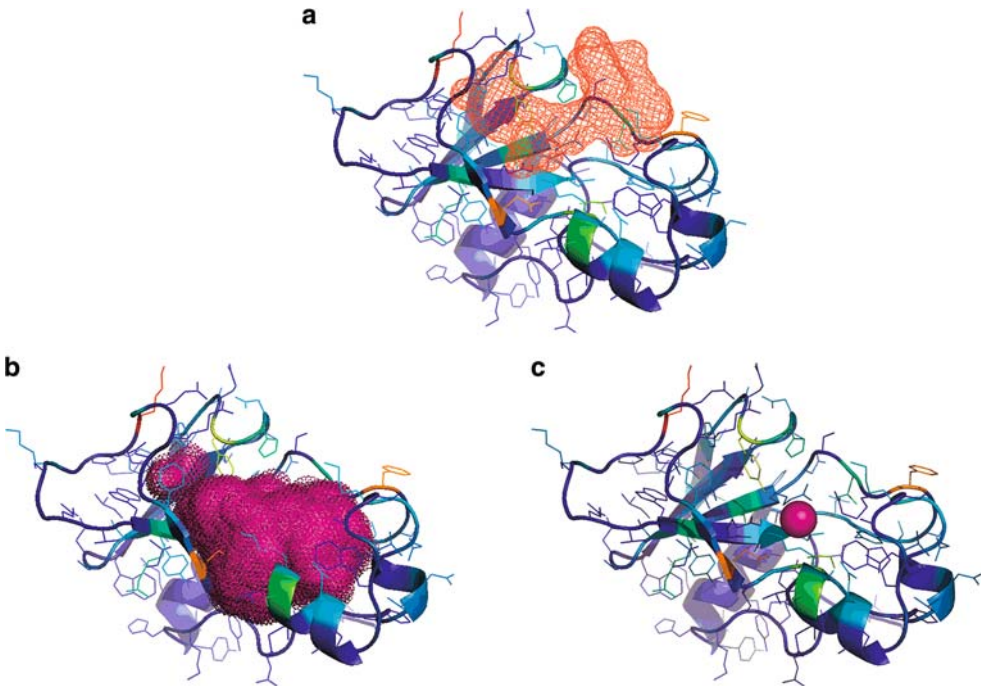

Fig. 8 **a** Experimental CIS for d(CGAC) mapped onto the NMR solution structure of unbound barnase. The surface is coloured according to the magnitude of the CIS values. The largest absolute changes are represented in red, moderate changes in yellow and green and smallest changes in cyan. The volume occupied by the ligand in the X-ray crystal structure of the complex barnase-d(CGAC) is shown as a red mesh. **b** *j*-surface (magenta) representing potential locations for the centre of the ligand. **c** Centre of the *j*-surface

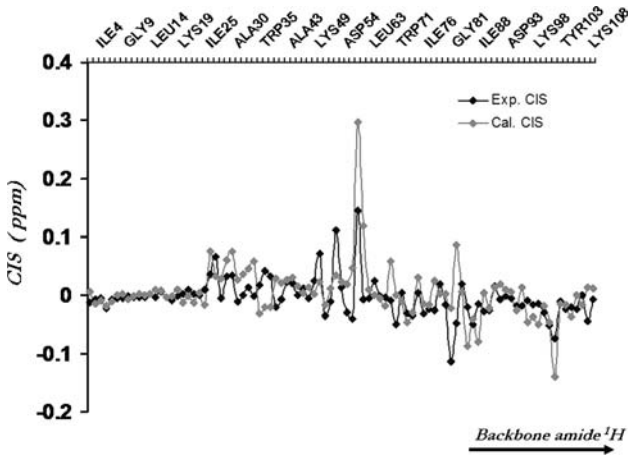

Fig. 9 Comparison between calculated and experimental CIS values of the protein backbone amide protons for the barnase-d(CGAC) complex

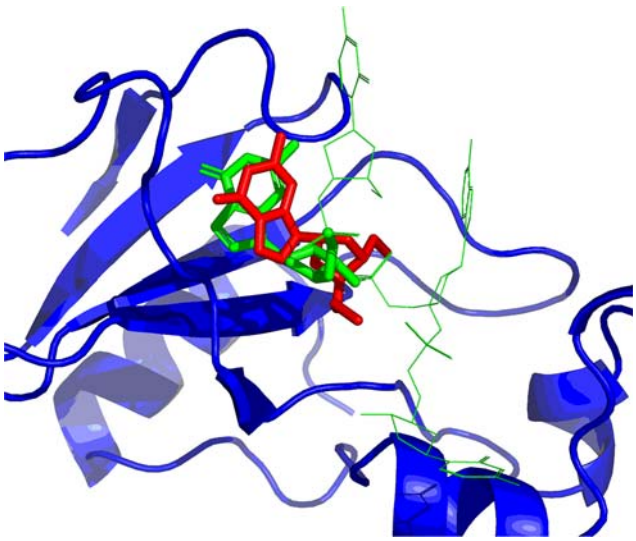

Fig. 10 Superposition of the guanine in the optimised CIS structure of the barnase-d(CGAC) complex (green) with the guanine in the X-ray crystal structure of the barnase-d(CGAC) complex (red). Thinner green lines show the rest of the d(CGAC)

identified by *Jsurf* to be inaccurate. This is the reason for the second stage of the calculations, where this initial location is used as a restraint for the generation of a large number of sterically allowed poses within *GOLD*: effectively, it does not matter if the initial location is inaccurate as long as it is not grossly wrong, since the ligand is allowed to move within *GOLD*. Our calculations so far on a range of proteins suggest that the *Jsurf* search is robust with respect to larger ligands (for example, containing several aromatic rings), but can fail if there are large chemical shift changes in the protein that are not due to direct interactions with the ligand. This point is a subject of ongoing research. At the second stage, the main requirement is to generate a large number of poses that are as diverse as possible. The final stage in *Shifty* acts as both a filter for the poses that best match the experimental shifts, and a refinement of the structure.

The method described here is a reasonably rapid search. The initial search is based solely on ring-current effects, and is therefore unable to place the ligand if it does not contain aromatic rings. By contrast, the final stage incorporates a more complete chemical shift calculation, and is thus able to provide some information on the location of

aliphatic parts of the molecule. However, our earlier study suggests that the locations of aromatic regions are better defined, not least because they tend to give rise to the largest intermolecular shift changes.

Our results show that in both cases the ligand binds in the active site, with the guanine located in the G-recognition binding site. The structure obtained from the CIS analysis is very different from the X-ray crystal structure reported for the barnase-d(GC) complex, where the guanine was observed in a secondary subsite. This is clearly not the most populated complex in solution and suggests that, in the X-ray crystal structure, binding is significantly perturbed by crystal packing interactions.

Acknowledgements We would like to thank AstraZeneca and the Biotechnology and Biological Sciences Research Council (grant BB/D015308/1) for funding; and Dr Arthur Moir for mass spectrometry and protein sequencing.

References

Baudet S, Janin J (1991) Crystal structure of a barnase-g(GpC) complex at 1.9 Å resolution. J Mol Biol 219:123–132

Buckle AM, Fersht AR (1994) Subsite binding in an RNAse: structure of a barnase tetranucleotide complex at 1.76 Å resolution. Biochemistry 33:1644–1653

Bycroft M, Ludvigsen S, Fersht AR, Poulsen FM (1991) Determination of the 3-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. Biochemistry 30:8697–8701

Cioffi M, Hunter CA, Packer MJ, Spitaleri A (2008a) Determination of protein-ligand binding modes using complexation-induced changes in ^{1}H NMR chemical shift. J Med Chem 51:2512–2517

Cioffi M, Hunter CA, Packer MJ (2008b) Influence of conformational flexibility on complexation-induced changes in chemical shift in a neocarzinostatin protein–ligand complex. J Med Chem 51:4488–4495

Day AG, Parsonage D, Ebel S, Brown T, Fersht AR (1992) Barnase has subsites that give rise to large rate enhancements. Biochemistry 31:6390–6395

Dobrodumov A, Gronenborn AM (2003) Filtering and selection of structural models: combining docking and NMR. Proteins: Struct Funct Genet 53:18–32

Dominguez C, Boelens R, Bonvin A (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125:1731–1737

Hunter CA, Packer MJ (1999) Complexation-induced changes in ^{1}H NMR chemical shift for supramolecular structure determination. Chem Eur J 5:1891–1897

Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748

Kohlbacher O, Burchardt A, Moll A, Hildebrandt A, Bayer P, Lenhof HP (2001) Structure prediction of protein complexes by an NMR-based protein docking algorithm. J Biomol NMR 20:15–21

Korzhnev DM, Bocharov EV, Zhuravlyova AV, Tischenko EV, Reibarkh MY, Ermolyuk YS, Schulga AA, Kirpichnikov MP, Billeter M, Arseniev AS (2001) ^{1}H, ^{13}C and ^{15}N resonance assignment for barnase. Appl Magn Reson 21:195–201

McCoy MA, Wyss DF (2002) Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. J Am Chem Soc 124:11758–11763

Meiering EM, Bycroft M, Lubienski MJ, Fersht AR (1993) Structure and dynamics of barnase complexed with 3′-GMP studied by NMR spectroscopy. Biochemistry 32:10975–10987

Morelli X, Dolla A, Czjzek M, Palma PN, Blasco F, Krippahl L, Moura JJG, Guerlesquin F (2000) Heteronuclear NMR and soft docking: an experimental approach for a structural model of the cytochrome c_{553}-ferredoxin complex. Biochemistry 39:2530–2537

Morelli XJ, Palma PN, Guerlesquin F, Rigby AC (2001) A novel approach for assessing macromolecular complexes combining soft-docking calculations with NMR data. Protein Sci 10:2131–2137

Mossakowska DE, Nyberg K, Fersht AR (1989) Kinetic characterization of the recombinant ribonuclease from *Bacillus amyloliquefaciens* (barnase) and investigation of key residues in catalysis by site-directed mutagenesis. Biochemistry 28:3843–3850

Palma PN, Krippahl L, Wampler JE, Moura JJG (2000) BiGGER: a new (soft) docking algorithm for predicting protein interactions. Proteins: Struct Funct Genet 39:372–384

Schägger H, von Jägow G (1991) Blue native electrophoresis for isolation of membrane–protein complexes in enzymatically active form. Anal Biochem 199:223–231

Schieborr U, Vogtherr M, Elshorst B, Betz M, Grimme S, Pescatore B, Langer T, Saxena K, Schwalbe H (2005) How much NMR data is required to determine a protein–ligand complex structure? Chembiochem 6:1891–1898

Stark J, Powers R (2008) Rapid protein-ligand costructures using chemical shift perturbations. J Am Chem Soc 130:535–545

van Dijk ADJ, Boelens R, Bonvin AMJJ (2005) Data-driven docking for the study of biomolecular complexes. FEBS J 272:293–312

Vinter JG (1996) Extended electron distributions applied to the molecular mechanics of some intermolecular interactions. 2. Organic complexes. J Comp Aided Mol Des 10:417–426