Journal of Biomolecular NMR, **14:** 115–122, 1999.
KLUWER/ESCOM

Validation of a new restraint docking method for solution structure determinations of protein–ligand complexes

Vladimir I. Polshakov[a,b], William D. Morgan[a], Berry Birdsall[a] & James Feeney[a,*]
[a]*Division of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K.;* [b]*Center for Drug Chemistry, Moscow 119815, Russia*

Received 4 January 1999; Accepted 31 March 1999

Key words: dihydrofolate reductase, ligand docking, protein–ligand complexes, structure determination

Abstract

A new method is proposed for docking ligands into proteins in cases where an NMR-determined solution structure of a related complex is available. The method uses a set of experimentally determined values for protein–ligand, ligand–ligand, and protein–protein restraints for residues in or near to the binding site, combined with a set of protein–protein restraints involving all the other residues which is taken from the list of restraints previously used to generate the reference structure of a related complex. This approach differs from ordinary docking methods where the calculation uses fixed atomic coordinates from the reference structure rather than the restraints used to determine the reference structure. The binding site residues influenced by replacing the reference ligand by the new ligand were determined by monitoring differences in ^{1}H chemical shifts. The method has been validated by showing the excellent agreement between structures of *L. casei* dihydrofolate reductase.trimetrexate calculated by conventional methods using a full experimentally determined set of restraints and those using this new restraint docking method based on an *L. casei* dihydrofolate reductase.methotrexate reference structure.

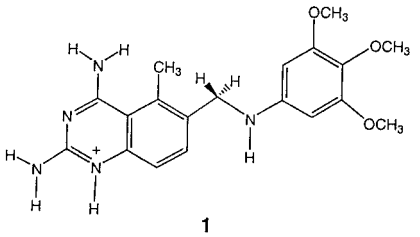
Abbreviations: DHFR, dihydrofolate reductase; MTX, methotrexate; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; rmsd, root mean square deviation; TMQ, trimetrexate.

Introduction

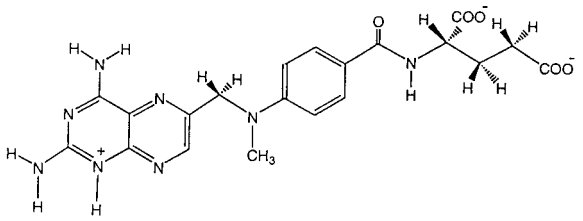
The structure determination of a protein–ligand complex in solution using multidimensional NMR and isotopically labelled samples has now become a fairly routine process. However, such determinations are still very time-consuming, particularly if one wants to obtain the high quality structures required for making meaningful comparisons between closely related complexes. It would be very useful to have NMR methods which could provide high quality structural information more efficiently and conveniently for such a series of related complexes involving the same protein. Several rapid structure determinations based on using docking methods have already been reported. These methods use an extensive set of protein–ligand NOEs to dock the ligand into a reference protein structure taken from crystal structure data on a related complex (Bennion et al., 1992; Fesik et al., 1992; Weber et al., 1992, 1993, 1994; Lian et al., 1994; Martorell et al., 1994; Byeon et al., 1995; Morgan et al., 1995; Gradwell et al., 1996; Johnson et al., 1997). In such calculations, large parts of the protein reference structure are held fixed while other parts around the binding site are allowed to move (Morgan et al., 1995) An obvious extension of this docking approach would be to use an NMR-determined solution structure as the reference structure. In contrast to the X-ray reference structure, one would now be dealing with the family of structures found in solution. In this paper we propose a variation of this method. Rather than using the atomic co-ordinates of the reference structures in the docking, this alternative approach uses the measured NMR restraints from

*To whom correspondence should be addressed.
E-mail: jfeeney@nimr.mrc.ac.uk

which the reference structures were generated. This paper describes such a method for determining the structure of a new protein–ligand complex based on using a composite list of restraints comprising a set of experimentally determined protein–ligand, ligand–ligand and protein–protein NMR restraints involving residues in and around the binding site for the new complex, and an additional set of protein–protein restraints, involving all remaining residues, extracted from a list previously used for determining the high resolution solution structure of a related complex. We have validated the method by comparing structures of the complex of *L. casei* DHFR and trimetrexate (**1**) determined using this approach with structures of the same complex obtained using the conventional structure determination method (Polshakov et al., 1999).



1



2

Restraint docking method

The full structure determinations of the complexes of *L. casei* dihydrofolate reductase with methotrexate (**2**) (Gargaro et al., 1998; Brookhaven Protein Data Bank 1ao8) and trimetrexate (Polshakov et al., 1999; Brookhaven PDB 1bzf) have been described earlier and here we illustrate the application of the restraint docking method using the data already available for these complexes. We choose the methotrexate complex as the reference structure and demonstrate how it is used for the restraint docking of trimetrexate to *L. casei* DHFR. The general protocol of this restraint docking method is shown in Figure 1 and includes

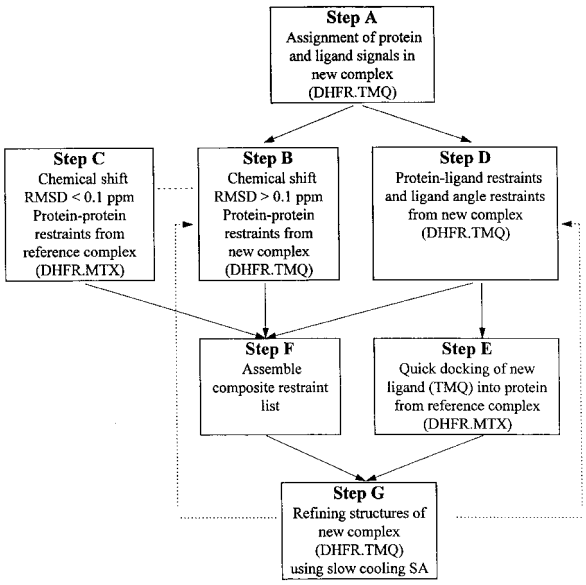


Figure 1. Protocol for the restraint docking method.

two main parts, assembling of the NMR restraints and carrying out the structure calculations.

Assembling the restraints

The first step (step A) is to make the resonance assignments for the protein and the bound ligand for the new complex (DHFR.trimetrexate in this case). In the present study, aimed at validating the method, these assignments were taken from the available data already published for the DHFR complex with trimetrexate (Polshakov et al., 1999). In fact, signal assignment procedures can often be simplified once the assignments for several well-determined related complexes have been made (Martorell et al., 1994; Polshakov et al., 1999): in some cases full assignments can be obtained without the need for recording the complete range of heteronuclear multidimensional experiments. Step B identifies those protein residues substantially influenced by changing the ligand and determines the experimental restraints involving the nuclei in the affected residues. For this, we compare the reported chemical shifts of the DHFR.trimetrexate and DHFR.methotrexate complexes (Soteriou et al., 1993; Gargaro et al., 1998; Polshakov et al., 1999), using the program NMRTABLE written in house, and identify the protein residues showing substantial chemical shift differences. In the case of a stereospecifically unassigned diastereotopic pair of protons, both combinations of the chemical shift differences were measured and the minimum values were

selected for the rmsd calculations. All the available ^{1}H chemical shifts for the complexes were compared for each residue and the rmsd values were calculated. Twenty-three residues had rmsd values $>$ 0.1 ppm and these were considered to be substantially influenced by replacement of the ligand: the chemical shift data for these residues are shown in Figure 2A. In the restraint docking method, all restraints (NOE-based distances, torsion angles and hydrogen bonds) involving these 23 residues need to be determined for the new complex. For the calculations described here, these restraints were already available because they had been determined for the conventional structure calculations of the DHFR.trimetrexate complex (Polshakov et al., 1999). In step C, the experimental restraints pertaining to the remaining 139 residues (whose chemical shifts were not significantly influenced by the change in ligand) were extracted from the list of restraints used previously for the reference structure (DHFR.methotrexate; Gargaro et al., 1998).

All docking procedures require an extensive list of restraints between the ligand and protein and in the next step D, the protein–ligand NOEs are assigned for the new complex (DHFR.trimetrexate) in order to obtain the corresponding distance restraints. Ligand torsion angle restraints are also included at this stage. For the calculations described here, the restraints involving the ligand were taken from the corresponding list used for the DHFR.trimetrexate full structure calculation (Polshakov et al., 1999). These constraints were used in two ways: firstly (step E), to determine an initial docked structure of trimetrexate with the reference structure (see *Structure calculations*) and secondly, to form part of the composite list of restraints (step F) used in the determination of the final structure (step G).

The composite list of restraints (B, C and D) used for the structure calculations of the DHFR.trimetrexate complex (see Table 1) contained 2787 NOEs (protein–protein: 960 (long range), 450 (medium range), 606 (sequential), 560 (intraresidue); protein–ligand: 75; ligand–ligand: 12). Eight ambiguous NOEs and 116 H-bonds were also included in the calculations. No intraresidue NOEs between atoms separated by three bonds (e.g. HN-Hα, Hα-Hβ) were included in the X-PLOR calculations because these (about 430 NOEs) had already been used for defining the torsion angles in the AngleSearch calculations (Polshakov et al., 1995).

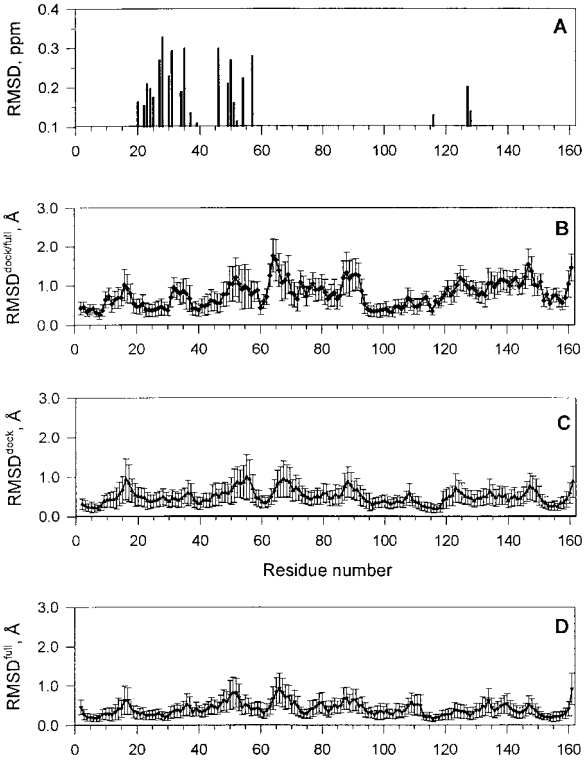


Figure 2. Plots of calculated rmsd parameters for the DHFR.trimetrexate complex plotted against residue number. (A) rmsd values of the ^{1}H chemical shift differences between the DHFR.trimetrexate and DHFR.methotrexate complexes: only residues with rmsd $>$ 0.1 ppm are shown. rmsd $= (\sum \Delta\delta^2/N)^{1/2}$ where $\Delta\delta$ are the chemical shift differences for the individual protons and N is the number of protons compared for each residue. (B) rmsd values ($\text{rmsd}^{\text{dock/full}}$) for the displacements over backbone atoms (Cα, C, N) calculated from the pairwise superimposition of each member of the family of DHFR.trimetrexate structures obtained from the restraint docking method on each member of the family of structures from the full conventional structure determination. (C) rmsd values ($\text{rmsd}^{\text{dock}}$) for the backbone atoms calculated from the pairwise superimposition on backbone atoms of residues 1–160 within a family of DHFR.trimetrexate structures calculated using the restraint docking method. (D) rmsd values ($\text{rmsd}^{\text{full}}$) for the backbone atoms calculated from the pairwise superimposition on backbone atoms of residues 1–160 within a family of DHFR.trimetrexate structures calculated using the full conventional method.

Structure calculations

The structure calculations were performed using X-PLOR 3.1 (Brünger, 1992) on Silicon Graphics (Origin 200 and O2) workstations. In step E, initial structures for DHFR.trimetrexate were generated from the family of 21 NMR structures of the DHFR.methotrexate complex determined earlier (Gargaro et al., 1998). At this stage, the methotrexate was removed from the structures and the new ligand, trime-

Table 1. NMR restraints and structural statistics for the DHFR.trimetrexate complex calculated using the restraint docking method

A. Restraints used in the final structure calculation		
NOEs	Full list	TMQ list[a]
Total	2787	565
Long range ($\|i-j\|>4$)	960	185
Medium range ($1<\|i-j\|\leq 4$)	450	96
Sequential ($\|i-j\|=1$)	606	118
Intraresidue	560	63
Ambiguous	8	8
Protein–ligand	75	75
Ligand–ligand	12	12
H-bonds	116	8
Dihedral angles		
Total	345	47
Phi (ϕ)	123	16
Psi (ψ)	120	14
Chi-1 (χ_1)	96	13
Chi-2 (χ_2)	4	2
TMQ	2	2
B. Constraint violations in the final ensemble of 19 structures		
Number of NOE constraint violations above 0.2 Å	0	
Number of dihedral angle violations above 3°	0	
XPLOR energies (kcal mol^{-1})[b]		
E_{NOE}	0.245 ± 0.122[c]	
E_{CDIH}	0.496 ± 0.047	
C. Deviations from idealised geometry		
Bonds (Å) $\times 10^3$	4.610 ± 0.013	
Angles (°)	0.326 ± 0.003	
Impropers (°)	0.144 ± 0.003	
D. Structural statistics for the final ensemble of 22 structures		
PROCHECK analysis		
% of residues in most favourable region of Ramachandran plot	80.2	
% of residues in non-allowed region of Ramachandran plot	0	
E. Pairwise superimposition (Å)		
Backbone (C, Cα, N) rmsd of residues 1–162	0.568 ± 0.102	
Heavy-atom rmsd of residues 1–162	1.129 ± 0.114	

[a]The TMQ list includes NMR restraints from the ligand and the 23 protein residues having large rmsd values of ^{1}H chemical shifts (see text and the caption of Figure 2A).
[b]The force constant used to calculate E_{NOE} was 50 kcal mol^{-1}Å^2. The force constant used to calculate E_{CDIH} was 200 kcal $mol^{-1}rad^{-2}$.
[c]Here and below: mean ± standard deviation.

trexate, was placed outside the protein (about 20 Å away). The initial docking of the trimetrexate into DHFR was achieved using only the protein–ligand NOEs for the DHFR.trimetrexate complex with 500 steps of Powell energy minimisation. During this initial stage, atoms of the protein residues were kept fixed, the vdW force constants were scaled down by a factor of 10^{-3}, and the force constants for angles and bonds were each increased by a factor of 10. Then all force constants were returned back to their standard values and only that for bonds was increased by a factor of 5. This was followed by another 500 steps of conjugate gradient energy minimisation. After this

stage, the ligand occupied roughly the correct position in its protein binding site.

All 21 structures were then subjected to the next stage of structure refinement using the composite list of experimental restraints (step G). The refinement protocol was essentially the same as that described earlier for the DHFR.trimetrexate structure determination (a slow cooling simulated annealing from 1000 K to 100 K for 20 ps followed by 1000 steps of conjugate gradient energy minimisation; Polshakov et al., 1999). Bond lengths were constrained during all dynamic stages of the slow cooling refinement using SHAKE (Ryckaert et al., 1977). Pseudoatom center averaging was applied for methyl groups and all distance constraints based on NOEs involving stereospecifically assigned methyl groups had their upper limits increased by 1 Å. An r^{-6} sum averaging, where the r^{-6} distance was weighted by the number of ambiguous NOEs (Nilges, 1995), was applied for all non-stereospecifically assigned methyl groups and methylene protons, and for the Hδ and Hϵ protons in phenylalanine and tyrosine rings. An r^{-6} sum averaging was used for all NOE restraints involving the 23 protein residues with substantial chemical shift rmsds between the trial and reference complexes (DHFR.trimetrexate and DHFR.methotrexate).

The final 19 structures were selected as those having no distance violation greater than 0.2 Å and no dihedral angle violations greater than 3°, and these structures were examined in the final analysis and comparison. Analysis of the final structures using PROCHECK-NMR/AQUA (Laskowski et al., 1996) indicated that there are no residues in the disallowed regions of the Ramachandran plot (see Table 1).

Analysis of calculated structures

Structures were visualised using INSIGHT II (Molecular Simulations Inc.) and analysed using X-PLOR 3.1 (Brünger, 1992).

In order to compare the families of the restraint docked structures with those from the previously determined solution structure of the DHFR.trimetrexate complex it is necessary to calculate the rmsd values for the pairwise superimposition between the two families A and B ($rmsd^{AB}$) (see Figure 2B). The $rmsd_i^{AB}$ values for each residue i were calculated using cross-pairwise fitting of each structure in family A from the restraint docking method on each structure from the earlier NMR determined family B for the DHFR.trimetrexate complex (418 pairs).

To assess the contributions to the $rmsd_i^{AB}$ from the differences in precision in the individual families, we have also calculated the pairwise rmsds for each residue i for each individual family, $rmsd_i^{A}$ and $rmsd_i^{B}$(see Figures 2C and 2D).

Results

Figure 3 shows a stereoview of the two mean structures of the DHFR.trimetrexate complex calculated by the restraint docking method and by the earlier conventional structure determination. One can see that there is very good agreement throughout the protein structure with particularly good overlap in the binding site environment. The observed perfect overlap of the ligand atoms in the two structures is especially noteworthy since the ligand atoms were not used for the superimposition of the two mean structures and no energy minimisation was carried out on the mean structures. When the two mean structures were superimposed on the backbone atoms (Cα, C, N) of residues 1–160, the rmsd value for the backbone atoms is 0.70 Å, the rmsd value for heavy atoms of the residues in the binding site is 0.52 Å, and for the ligand heavy atoms it is 0.17 Å. Obviously one would expect the largest differences between the two structures to arise in those regions remote from the binding site where the constraints used for the validation of the method are taken from two different complexes. These differences in structure give some indication of the small errors that could arise in these regions when using this method.

Figure 4 illustrates the precision of the two calculated families and the good agreement between them. Figure 4A shows all protein backbone and ligand heavy atoms for the two families of calculated structures. For most regions of the protein, the agreement between the two families is within the precision seen for these regions in each family: this is particularly true for the residues within the binding site. The latter is seen clearly in Figure 4B, which shows the superimposed families for the binding site region where the set of atoms used for superimposition also included side-chain and ligand heavy atoms. From a detailed comparison of the rmsd values given in Figure 2 it is seen that the regions of the protein showing the largest values are the same for both families. The same regions also show the largest values in the cross comparison between the two families ($rmsd_i^{AB}$). The most significant differences in all cases are seen for the loop

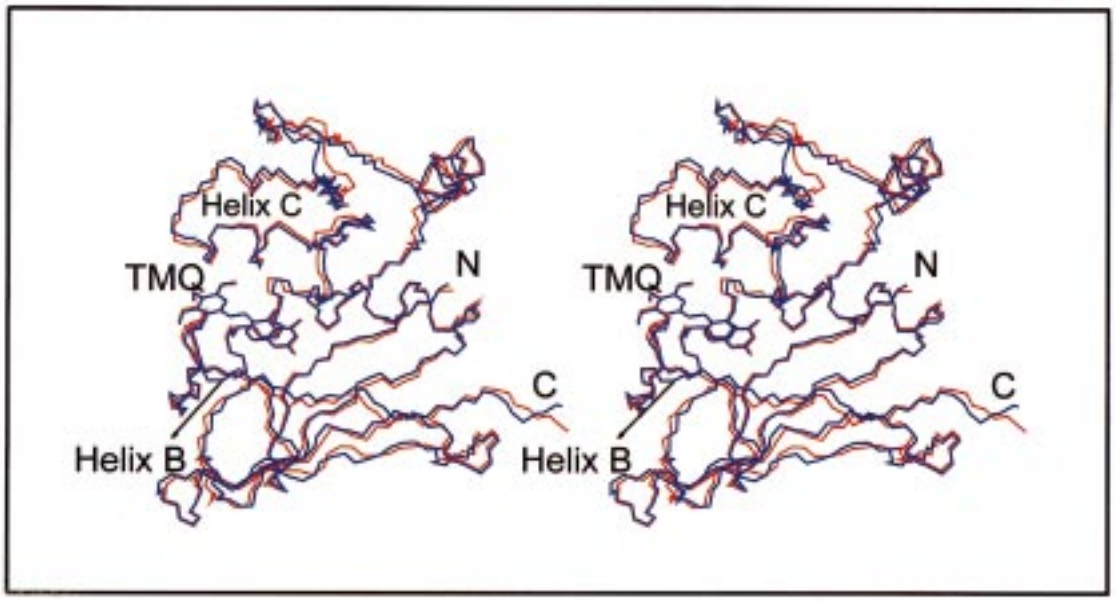


Figure 3. Stereoview of the two mean structures calculated from the families of structures obtained by the restraint docking method (blue) and by the conventional method (red) and superimposed on the backbone atoms of residues 1–160. Only the backbone atoms of the protein and the heavy atoms of the ligand are shown.

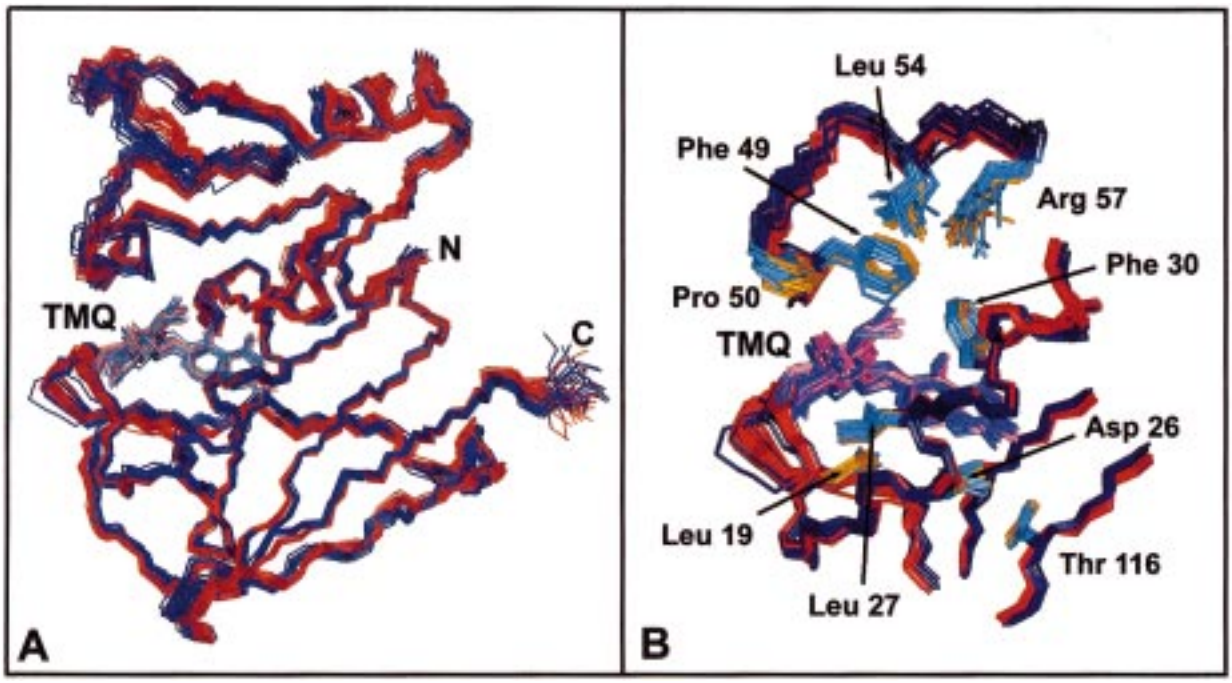


Figure 4. Superimposition of two families of DHFR.trimetrexate structures calculated using the restraint docking method (blue) and the full conventional method (red). The ligand in pink is from the restraint docking method and the ligand in light blue is from the conventional method. (A) This superimposition was performed on the backbone atoms of residues 1–160. Only the backbone atoms of the protein and the heavy atoms of the ligand are shown. (B) This superimposition was performed using the heavy atoms of the residues in the binding site. The side-chains of the protein residues from structures calculated using the restraint docking method are shown in yellow and those from the conventional structural determination in light blue. Only the heavy atoms of the ligand and the protein residues in the binding site are shown.

regions only. However, even in these regions the differences are generally small ($rmsd^A$ and $rmsd^B$ are less than 1.0 Å and $rmsd_i^{AB}$ is less than 1.5 Å).

Discussion and conclusions

Choosing which NOEs need to be measured for the new complex

While all the protein–ligand NOEs must be measured for this restraint docking method, the only protein–protein NOEs which need to be determined experimentally are those involving residues influenced by replacement of the ligand. We have adopted the approach of using ^{1}H chemical shift differences to indicate which parts of the structure of a protein–ligand complex have been influenced by replacing one ligand with another. The ^{1}H chemical shifts are particularly sensitive to any changes involving aromatic rings in such complexes. In this study we use the absence of substantial chemical shift differences between residues in two protein–ligand complexes as a criterion for determining the regions of the proteins that are essentially structurally identical. Figure 2A indicates the 23 (of the 162) DHFR residues which have large chemical shift differences (rmsd $>$ 0.1 ppm for ^{1}H of all protons in the residue) between the complexes with methotrexate and trimetrexate. The remaining 139 residues with only very small chemical shift differences are assumed to be from regions of the protein where the structure is essentially identical. Based on this, the NMR experimental constraints (NOEs, torsion angles and hydrogen bonds) involving the residues from this unaffected group of residues can be extracted from a list of constraints used previously for structure determination of the reference structure (DHFR.methotrexate complex (Gargaro et al., 1998)). These are included in the final composite list of restraints used in the docking of the DHFR.trimetrexate complex. Constraints involving the 23 protein residues with chemical shift rmsd $>$ 0.1 ppm and those involving the ligand itself were determined from 2D and 3D NOESY-based spectra of the DHFR.trimetrexate complex. In the final structures, these 23 affected residues either occupy or are close to the ligand binding sites. Several additional protein residues, specifically those close to one of the 23 residues or those close to the ligand, will of necessity also be involved in the measured NOEs. The good agreement observed between the restraint docked structure and the conventional structure calculated for the DHFR.trimetrexate complex validates the criteria we use for selecting which NOEs need to be experimentally determined and which NOEs can be taken directly from the restraint list used for generating the reference structure.

In the present example where the method is being demonstrated, the final list of constraints was assembled from the detailed data obtained during a full independent structural determination. In the case of a new complex being examined de novo, it would be necessary, of course, to introduce iterative steps during the structure refinement stage (involving step G with steps B and D) to assist in assigning the protein–protein and protein–ligand NOEs.

Advantages of the method

The good agreement between the structures obtained using this restraint docking method and those from the conventional structure determination validate the use of the restraint docking approach for obtaining high quality structures. This is particularly true for the bound ligand and the binding site region, although the overall structures are also in very good agreement. This method leads to time-saving in two important ways. Firstly, one needs to measure fewer NOEs and other restraints, and secondly, the structure refinement is considerably simplified particularly in dealing with violated and ambiguous NOEs. Any misassignments or incorrect calibration of NOEs can usually be detected very readily when these involve protons within regions of the protein which are usually well determined from other measurements. The iterative procedure used in the full structure calculations to ensure that all these NOEs are consistent with the structure and free of assignment errors is a lengthy process. Obtaining the assignments and measuring the NOEs for protons in the bound ligand and the 'chemical shift affected' protein residues in the new protein–ligand complex will be straightforward in cases where well documented and fully assigned spectra of related complexes are already available.

Understanding structure–function relationships often involves comparing structures of a series of protein complexes formed by binding different ligands to the same protein. NMR has already proved itself to be a useful albeit time-consuming method for determining such structures in solution. Methods such as the one described here will allow for more rapid determination of structures of a series of related complexes in the future.

Acknowledgements

V.I.P. acknowledges the award of a Howard Hughes Medical Institute International Scholarship and a grant from the Russian Foundation for Basic Research (97-04-48245). We thank Dr. David Neuhaus for his advice and useful discussions which influenced the development of this restraint docking method.

References

Bennion, C., Connolly, S., Gensmantel, N.P., Hallam, C., Jackson, C.G., Primrose, W.U., Roberts, G.C.K., Robinson, D.H. and Slaich, P.K. (1992) *J. Med. Chem.*, **35**, 2939–2951.

Brünger, A.T. (1992) *X-PLOR 3.1: A system for X-ray crystallography and NMR*, Yale University Press, New Haven, CT.

Byeon, I.L., Kelley, R.F., Mulkerrin, M.G., An, S.S.A. and Llinas, M. (1995) *Biochemistry*, **34**, 2739–2750.

Fesik, S.W., Neri, P., Meadows, R., Olejniczak, E.T. and Gemmecker, G. (1992) *J. Am. Chem. Soc.*, **114**, 3165–3166.

Gargaro, A.R., Soteriou, A., Frenkiel, T.A., Bauer, C.J., Birdsall, B., Polshakov, V.I., Barsukov, I.L., Roberts, G.C.K. and Feeney, J. (1998) *J. Mol. Biol.*, **277**, 119–134.

Gradwell, M.J. and Feeney, J. (1996) *J. Biomol. NMR*, **7**, 48–58.

Johnson, J.M., Meiering, E.M., Wright, J.E., Pardo, J., Rosowsky, A. and Wagner, G. (1997) *Biochemistry*, **36**, 4399–4411.

Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) *J. Biomol. NMR*, **8**, 477–486.

Lian, L.Y., Barsukov, I.L., Sutcliffe, M.J., Sze, K.H. and Roberts, G.C.K. (1994) *Methods Enzymol.*, **239**, 657–700.

Martorell, G., Gradwell, M.J., Birdsall, B., Bauer, C.J., Frenkiel, T.A., Cheung, H.T.A., Polshakov, V.I., Kuyper, L. and Feeney, J. (1994) *Biochemistry*, **33**, 12416–12426.

Morgan, W.D., Birdsall, B., Polshakov, V.I., Sali, D., Kompis, I. and Feeney, J. (1995) *Biochemistry*, **34**, 11690–11702.

Nilges, M. (1995) *J. Mol. Biol.*, **245**, 645–660.

Polshakov, V.I., Frenkiel, T.A., Birdsall, B., Soteriou, A. and Feeney, J. (1995) *J. Magn. Reson.*, **B108**, 31–43.

Polshakov, V.I., Birdsall, B., Frenkiel, T.A., Gargaro, A.M. and Feeney, J. (1999) *Protein Sci.*, **8**, 467–481.

Ryckaert, J.-P., Ciccotti, G. and Berendsen, H.J.C. (1977) *J. Comput. Phys*., **23**, 327–341.

Soteriou, A., Carr, M.D., Frenkiel, T.A., McCormick, J.E., Bauer, C.J., Sali, D., Birdsall, B. and Feeney, J. (1993) *J. Biomol. NMR*, **3**, 535–546.

Weber, D.J., Gittis, A.G., Mullen, G.P., Abeygunawardana, C., Lattman, E.E. and Milvan, A.S. (1992) *Proteins Struct. Funct. Genet.*, **13**, 275–287.

Weber, D.J., Libson, A.M., Gittis, A.G., Lebowitz, M.S. and Mildvan, A.S. (1994) *Biochemistry*, **33**, 8017–8028.

Weber, D.J., Serpersu, E.H., Gittis, A.G., Lattman, E.E. and Mildvan, A.S. (1993) *Proteins Struct. Funct. Genet.*, **17**, 20–35.