

Article

A simple and reliable approach to docking protein–protein complexes from very sparse NOE-derived intermolecular distance restraints

Chun Tang & G. Marius Clore*

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, 20892-0520, Maryland, USA

Received 18 June 2006; Accepted 7 July 2006

Key words: protein docking, sparse NOE data, reverse isotope labeling, deuteration, ‘reduced’ radius of gyration, rigid body minimization

Abstract

A simple and reliable approach for docking protein–protein complexes from very sparse NOE-derived intermolecular distance restraints (as few as three from a single point) in combination with a novel representation for an attractive potential between mapped interaction surfaces is described. Unambiguous assignments of very sparse intermolecular NOEs are obtained using a reverse labeling strategy in which one of the components is fully deuterated with the exception of selective protonation of the δ -methyl groups of isoleucine, while the other component is uniformly ^{13}C -labeled. This labeling strategy can be readily extended to selective protonation of Ala, Leu, Val or Met. The attractive potential is described by a ‘reduced’ radius of gyration potential applied *specifically* to a subset of interfacial residues (those with an accessible surface area $\geq 50\%$ in the free proteins) that have been delineated by chemical shift perturbation. Docking is achieved by rigid body minimization on the basis of a target function comprising the sparse NOE distance restraints, a van der Waals repulsion potential and the ‘reduced’ radius of gyration potential. The method is demonstrated for two protein–protein complexes (EIN–HPr and IIA^{Glc}–HPr) from the bacterial phosphotransferase system. In both cases, starting from 100 different random orientations of the X-ray structures of the free proteins, 100% convergence is achieved to a single cluster (with near identical atomic positions) with an overall backbone accuracy of ~ 2 Å. The approach described is not limited to NMR, since interfaces can also be mapped by alanine scanning mutagenesis, and sparse intermolecular distance restraints can be derived from double cycle mutagenesis, cross-linking combined with mass spectrometry, or fluorescence energy transfer.

Introduction

Most biological processes involve protein–protein interactions. Despite advances in methodology, *ab initio* docking of protein–protein complexes is still far from reliable (Russel et al., 2004; Mendez et al., 1995) and experimental structure determination of complexes presents significant technical challenges (Fahmy and Wagner, 2002; Bonvin

et al., 2005; van Dijk et al., 2005; Takeuchi et al., 2006). NMR is particularly suited for low affinity complexes since they are generally difficult to crystallize. The geometric information used to dock two proteins by NMR principally resides in short (<6 Å) intermolecular NOE-derived interproton distances (Wüthrich, 1986; Clore and Gronenborn 1998). In favorable cases (Garrett et al., 1999; Wang et al., 2000; Williams et al., 2004, 2005), residual dipolar couplings (RDCs) measured in weakly aligned samples can yield

*To whom correspondence should be addressed. E-mail: mariusc@intra.niddk.nih.gov

orientational information (Bax and Grishaev, 2005). Although powerful, RDCs can be difficult to apply either because complete saturation cannot be achieved or because the alignment media preferentially interact with one of the components (Cai et al., 2003; Suh et al., 2006).

It has recently been shown that reliable docking can be achieved using RDC data in combination with highly ambiguous intermolecular distance restraints (HADR), each comprising between 400 and 3000 individual distances represented by a $\Sigma(r^{-6})^{-1/6}$ sum, that provide an attractive potential between the interaction surfaces (Clare and Schwieters, 2003). The latter can be delineated by chemical shift perturbation, cross-saturation, solvent paramagnetic resonance effects, or alanine scanning mutagenesis (Bonvin et al., 2005; Takeuchi et al., 2006). The landscape of the HADR potential, however, is extremely rough and characterized by multiple local minima due to the fact that once one of the possible assignments is satisfied, $\Sigma(r^{-6})^{-1/6}$ summation effectively eliminates the atomic forces from all other possible assignments (Kuszewski et al., 2004). Nevertheless, HADRs without RDCs have been successfully used for docking (Dominguez et al., 2003), but more often than not it is difficult (without prior knowledge of the correct answer) to unambiguously and reliably differentiate between the correct solution and the vast majority of incorrect alternatives. This is especially the case when the coordinates of the free proteins are employed in the calculations instead of using starting coordinates obtained by simply prying the coordinates of the complex apart.

As the complexity of protein–protein complexes being analyzed by NMR increases, the number of unambiguous intermolecular NOEs that can be assigned *a priori* decreases. This renders the early stages of an NMR structure determination of a protein–protein complex particularly difficult and error prone, particularly when RDC data that accurately define the relative orientation of the two components are not available. In this paper we present a simple approach for obtaining reliable preliminary docking of protein–protein complexes based on an alternative representation of the attractive potential between mapped interaction surfaces in combination with very sparse intermolecular distance restraints (as few as three) derived from selective labeling

experiments. We illustrate the application of the approach to two complexes of the bacterial phosphotransferase system, namely the 40 kDa EIN–HPr complex and the 30 kDa IIA^{Glc}–HPr complexes, whose structures have been previously solved by NMR (Garrett et al., 1999; Wang et al., 2000).

Materials and methods

Preparation of 3,3-D₂- α -ketobutyrate

Sodium α -ketobutyrate was purchased from Sigma-Aldrich (Product number K0875, St. Louis, MO). Exchange of the basic methylene protons to deuterons to make sodium 3,3-D₂- α -ketobutyrate was carried out by dissolving 50 mg sodium α -ketobutyrate at a concentration of 10 mM in D₂O preadjusted to pH* 10.5 with NaOD and incubating the solution at 42°C for 2 h. The H–D exchange process was monitored by the disappearance of methylene NMR resonance, and the resulting product only carries protons at the methyl group.

Protein expression and purification

The N-terminal domain of enzyme I (EIN) and HPr were expressed and purified as described previously (Garrett et al., 1999). U-[²H]/[¹H- δ -methyl-Ile]-EIN labeling was achieved by growing bacteria in D₂O M9 minimal medium with [²H₇, ¹²C₆]-glucose as the sole carbon source and by adding 50 mg of 3,3-D₂- α -ketobutyrate (as the biosynthetic precursor for Ile) per liter of medium 1 hour prior to induction with isopropyl β -D-1-thiogalactopyranoside. U-[¹³C]-HPr was obtained by growing the bacteria in minimal medium with ¹³C₆-glucose as the sole carbon source.

NMR spectroscopy

The sample for NMR spectroscopy comprised a 0.5 mM 1:1 complex of U-[¹³C,¹H]-HPr and U-[²H]/[¹H- δ -methyl-Ile]-EIN in 99.996% D₂O containing 10 mM sodium phosphate buffer, pH 7. Spectra were recorded at 40°C on a Bruker DRX600 spectrometer equipped with a z-shielded triple resonance cryoprobe. Assign-

ments were taken directly from Garrett et al. (1999). Intermolecular NOEs between protons attached to ^{13}C on HPr and the δ -methyl group of Ile residues were observed in a 3D ^{13}C -separated/ ^{12}C -filtered NOE spectrum recorded with a mixing time of 300 ms. The spectrum was processed and analyzed using the programs NmrPipe (Delaglio et al., 1995) and PIPP (Garrett et al., 1991), respectively.

Interface mapping

The criteria used to delineate the interfacial residues from $^1\text{H}_\text{N}/^{15}\text{N}$ chemical shift perturbation were described previously (Clare and Schwieters, 2003) and the complete list of interfacial residues delineated in this manner for the EIN–HPr and IIA^{Glc}–HPr complexes is given in Clare and Schwieters (2003). In the current application we only make use of interfacial residues whose solvent accessible surface area is greater than 50% of that in a Gly–X–Gly tripeptide segment, as calculated using the program GetArea (Fraczkiewicz and Braun, 1998). The target value for the ‘reduced’ radius of gyration (R_{gyr}) restraint is determined by calculating the R_{gyr} value for the interfacial residues of each of the two partner proteins, selecting the smaller of the two values and subtracting 1 Å from this value. This ensures that a force is always exerted by the R_{gyr} energy term (Kuszewski et al., 1999).

Starting coordinates for docking

The starting coordinates used for the docking calculations comprise the X-ray structures of free HPr (PDB code IPOH; Jia et al., 1993), EIN (PDB code 1ZYM; Liao et al., 1996) and IIA^{Glc} (PDB code 2F3G; Feese et al., 1997). The structures of EIN–HPr (PDB code 3EZA; Garrett et al., 1999) and IIA^{Glc}–HPr (PDB code 1GGR; Wang et al., 2000) were previously determined by NMR. The EIN–HPr complex was determined *de novo* (Garrett et al., 1999) on the basis of both NOE and residual dipolar couplings (RDCs) and hence there are backbone r.m.s. differences relative to the coordinates of the free crystal structures (~ 1 Å for EIN and ~ 0.6 Å for HPr); therefore the reference EIN–HPr structure used in this study was obtained by best-fitting the back-

bone coordinates of the free crystal structures of EIN (either molecule A or B of the asymmetric crystal unit) and HPr onto the corresponding NMR coordinates of the complex. The IIA^{Glc}–HPr complex (Wang et al., 2000) was determined by conjoined rigid body/torsion angle dynamics on the basis of intermolecular NOE data and RDCs with the backbone coordinates of IIA^{Glc} and HPr fixed to those of the free crystal structures (molecule A in the case of IIA^{Glc}). Hence, the published IIA^{Glc}–HPr structure was used as the reference structure in those calculations using the coordinates of molecule A of IIA^{Glc}. For those calculations using the coordinates molecule B of IIA^{Glc}, the reference structure was obtained by best-fitting the coordinates of molecule B of the free crystal structure to the IIA^{Glc} coordinates in the IIA^{Glc}–HPr complex. To generate the starting configurations for the docking calculations, the coordinates of EIN and IIA^{Glc} were fixed, and the orientation and positions of HPr relative to EIN and IIA^{Glc} were randomized to generate 100 starting structures each.

Docking protocol

Rigid body docking was performed using the internal coordinate module (IVM; Schwieters and Clare, 2001) of Xplor-NIH (Schwieters et al., 2003, 2006). The coordinates of EIN or IIA^{Glc} were held fixed, while those of HPr were free to rotate and translate. The target function comprises only three terms: a square-well quadratic term for the interproton distance restraints (E_{NOE}), a quartic van der Waals repulsion term (E_{repel} ; Nilges et al., 1988) and the ‘reduced’ radius of gyration term (E_{rgyr} ; Kuszewski et al., 1999). The force constants for E_{rgyr} and E_{NOE} are held constant throughout at values of 100 and 30 kcal mol⁻¹ Å⁻², respectively. The force constant for E_{repel} is geometrically increased over 116 cycles from 0.004 to 1 kcal mol⁻¹ Å⁻⁴, while simultaneously decreasing the van der Waals radius scale factor for 0.9 to 0.75. The force constant and van der Waals radius scale factor is then reset and the minimization process repeated. This is carried out two more times for a total of four complete cycles of ramping the adjustable parameters of E_{repel} , thereby ensuring highly converged structures. An Xplor-NIH

script for this protocol can be downloaded from <http://www.spin.niddk.nih.gov/clore>

Results and discussion

Selective observation of intermolecular NOEs involving the δ -methyl of Ile

Conventionally intermolecular NOEs between carbon-attached protons are obtained from a 3D ^{13}C -separated/ ^{12}C -filtered NOE experiment in which one partner is at natural isotope abundance and the other uniformly ^{13}C -labeled (Clore and Gronenborn, 1995). For larger systems, the unambiguous assignment of the proton attached to ^{12}C is often very difficult owing to extensive ^1H resonance overlap. To circumvent this problem, we substitute the protein at natural isotope abundance by one that is fully deuterated with the exception of selective protonation of the δ -methyl groups of Ile by growing the bacteria in D_2O minimal medium containing 3,3- D_2 - α -ketobutyrate as the biosynthetic precursor of Ile. This reverse labeling strategy draws analogy to schemes previously described for selective ^{13}C -labeling of certain residues (Goto et al., 1996; Tang et al., 2005; Yu et al., 2005). Because there are relatively few Ile residues in a protein and even fewer at the interface, unambiguous intermolecular NOE assignments can be readily obtained. The deuterated background also largely eliminates spin-diffusion within one of the partners, thereby enabling weak NOEs to be detected through the use of longer mixing times. A 2D projection of a 3D ^{13}C -separated/ ^{12}C -filtered NOE spectrum of a 40 kDa complex of U- ^{13}C , ^1H -HPr and U- ^{13}C , ^1H -EIN is shown in Figure 1, illustrating the observation of several NOEs originating from protons attached to ^{13}C on HPr to two protonated Ile δ -methyl groups attached to ^{12}C on EIN. This labeling strategy can be readily extended to selective protonation of Ala, Leu, Val or Met. [Note that isotope scrambling in the case of Ala can be circumvented by use of either an appropriate auxotrophic strain or by the addition of the transaminase inhibitor β -chloro-L-alanine (Kato et al, 1991; Mueller et al., 2003). In the latter case, the medium would contain a mixture of methyl-protonated Ala deuterated at the $\text{C}\alpha$ position together with the other 19 amino acids in fully deuterated form].

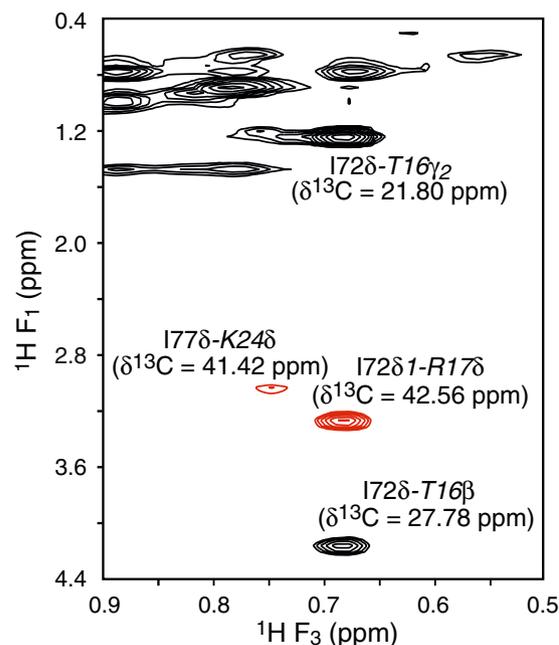


Figure 1. 2D $^1\text{H}(\text{F}_1)$ - $^1\text{H}(\text{F}_3)$ projection of a 3D ^{13}C -separated(F_2)/ ^{12}C -filtered(F_3) NOE spectrum (300-ms mixing time) recorded on a 40 kDa complex of U- ^{13}C , ^1H -labeled HPr and U- ^{13}C , ^1H -labeled EIN. Intermolecular NOE cross-peaks are labeled (HPr residues in *italics*) and the red contours indicate that the peak is folded in the $^{13}\text{C}(\text{F}_2)$ dimension. Unmarked peaks arise from incomplete suppression of diagonal auto-correlation signals.

The 'reduced' radius of gyration restraint

A very sparse set of intermolecular NOEs originating from a single point does not provide sufficient information content to uniquely dock a protein-protein complex. Thus, additional information is required in the form of an attractive potential between the two protein interfaces whose boundaries have been delineated, for example, by chemical shift perturbation. Rather than make use of a set of HADR whose corresponding pseudo-potential is characterized by a very rough energy hypersurface with multiple local minima (Kuszewski et al., 2004), we describe the attractive potential between the two protein interfaces by means of a single 'reduced' radius of gyration (R_{gyr}) potential term (E_{gyr} ; Kuszewski et al., 1999) applied *specifically* to the interfacial residues. The 'reduced' R_{gyr} is defined as the r.m.s. distance from each atom of the selected interfacial residues to their centroid. The energy hypersurface for E_{gyr} is smooth and, by applying this term solely to

interfacial residues, serves to optimize the complementarity of the two interfaces, thereby providing both translational and rotational information. Thus, the ‘reduced’ R_{gyr} term serves as a soft, stereochemically guided, intermolecular packing potential between two complementary interaction surfaces.

It should be noted that the interfacial residues in a complex make up a disk rather than a sphere. Although the R_{gyr} formalism assumes a spherical region, the use of rigid body docking (either full rigid body or conjoined rigid body/torsion angle dynamics where the interfacial side chains are given torsional degrees of freedom) together with the fact that interactions surfaces are small relative to the total surface of the proteins ensure that the ‘reduced’ R_{gyr} restraint does not introduce distortions but simply serves as a very simple attractive potential between two interfaces.

The criteria used to define the interface from chemical shift mapping are described in detail by Clore and Schwieters (2003). In this application, only solvent exposed residues (defined as having $\geq 50\%$ accessible surface area in the free proteins) are selected from the mapped interfacial residues to compute the E_{rgyr} term. Selection of this subset of residues is based on two observations: (a) chemical shift perturbation of the exposed interfacial residues are more likely to arise directly as a conse-

quence of complexation, and (b) the radial distribution of the exposed residues is thinner than that of all mapped interfacial residues, thereby making E_{rgyr} more sensitive to subtle changes in the relative positions of the two proteins. The ‘reduced’ R_{gyr} target value is obtained by subtracting 1 Å from the smaller of the two R_{gyr} values calculated for the selected interfacial residues of each of the two free proteins. This ensures that an attractive force is always present. It is important to stress that atomic overlap is prevented by the presence of a compensatory quartic van der Waals repulsion term (E_{repeI} ; Nilges et al., 1988).

Docking calculations with the ‘reduced’ R_{gyr} restraint and sparse intermolecular NOE data

The docking calculation, carried out using Xplor-NIH (Schwieters et al., 2003, 2006) makes use of only rigid body minimization (Clore, 2000) subject to a square-well potential for the sparse intermolecular NOE distance restraints (E_{NOE}), E_{rgyr} and E_{repeI} . Details of the protocol are provided in the Methods section. The method was tested using two complexes from the bacterial phosphotransferase system, EIN–HPr and IIA^{Glc}–HPr whose solution structures have been solved previously by conventional means and whose interfaces have been

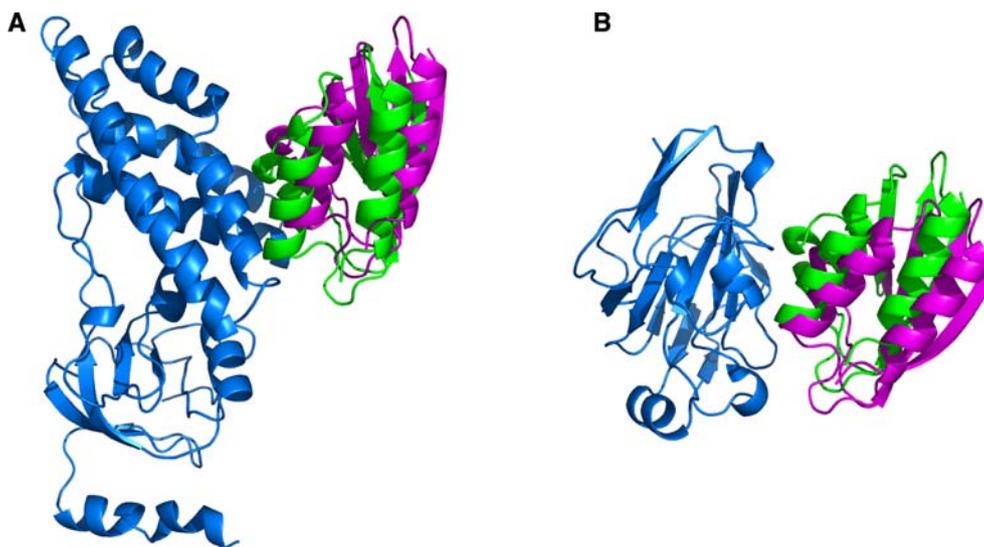


Figure 2. Comparison of the docked structures of the (A) EIN–HPr and (B) IIA^{Glc}–HPr complexes calculated on the basis of three intermolecular distances and the E_{rgyr} potential with the published structure based on a full complement of intermolecular NOEs and dipolar couplings. The backbone rms difference between the structures is 2.0 Å for EIN(molecule A)–HPr and 1.9 Å for IIA^{Glc}(molecule A)–HPr. EIN and IIA^{Glc}, blue; HPr docked, purple; HPr in published complexes green.

mapped by chemical shift perturbation (Garrett et al., 1999; Wang et al., 2000).

For the EIN–HPr complex, the experimental data comprising three intermolecular NOE distance restraints from Ile-72(C δ H $_3$) of EIN (Figure 1) were employed. These NOEs were all generously classified as weak with an upper bound of 5 Å plus an additional 0.5 Å for every methyl group involved (to account for the higher apparent intensity of methyl resonances; Clore et al., 1987), even though their actual distances in the published EIN–HPr complex (Garrett et al., 1999) are significantly shorter. For the IIA^{Glc}–HPr calculations, synthetic data were generated from the published structure and comprised three distance restraints from Ile-45 of IIA^{Glc}, with an upper bound of 5 Å plus an additional 0.5 Å for every methyl group involved. The restraints are represented by $\Sigma(r^{-6})^{-1/6}$ sums. A summary of the distance restraints employed is given in Table 1. It should be noted that all the distance restraints arise from a single point (namely the δ -methyl of an Ile residue) and therefore only provide a pivot point for the docking calculations.

The starting coordinates were the X-ray structures of the free proteins (Jia et al., 1993; Liao et al., 1996; Feese et al., 1997). The X-ray structures of free EIN (Liao et al., 1996) and IIA^{Glc} (Feese et al., 1997) have two molecules in their asymmetric units and calculations were carried out for each molecule to assess the impact of coordinate differences. While the overall r.m.s. difference for the backbone between the different molecules is generally

Table 1. Intermolecular distance restraints for docking calculations (Residues of HPr denoted in italics)

Restraint	$\Sigma(r^{-6})^{-1/6}$ distance in published structure of complexes (Å) ^a
EIN–HPr	
I72C δ H $_3$ – <i>T16Hβ</i>	3.8
I72C δ H $_3$ – <i>T16Cγ H$_3$</i>	3.3
I72C δ H $_3$ – <i>R17CδH$_2$</i>	2.7
IIA ^{Glc} –HPr	
I45C δ H $_3$ – <i>K24Cγ H$_2$</i>	4.3
I45C δ H $_3$ – <i>K24CδH$_2$</i>	3.3
I45C δ H $_3$ – <i>K24Cϵ H$_2$</i>	3.8

^aTaken from the coordinates of EIN–HPr (PDB accession code 3EZA; Garrett et al., 1999) and IIA^{Glc}–HPr (PDB accession code 1GGR; Wang et al., 2000).

small (0.3–0.4 Å), there is much greater variability in the coordinates of the interfacial sidechains (0.7–1.2 Å), as expected given their high degree of solvent accessibility (Table 2). The positions and orientations of HPr relative to EIN or IIA^{Glc} were randomized to generate 100 different starting positions (Figure 3). In every single case, the calculations converged to a single cluster with identical coordinate positions (within 0.001 Å). The backbone r.m.s. difference relative to the published structures was 2.0 and 2.1 Å for EIN–HPr and 1.7 and 1.9 Å for IIA^{Glc}–HPr, depending on the coordinates of EIN and IIA^{Glc} employed (Figure 2). The protocol works equally well for another complex,

Table 2. Atomic r.m.s. differences between the A and B chains of the crystal structures of free EIN and IIA^{Glc} ^a

	Atomic r.m.s. difference (Å)	
	EIN	IIA ^{Glc}
All backbone	0.51	0.24
All heavy atoms	0.97	0.77
Interfacial region backbone	0.32	0.39
Interfacial region heavy atoms	0.72	1.20

^aThe PDB codes for the crystal structures of free EIN and IIA^{Glc} are 1ZYM (Liao et al., 1996) and 2F3G (Feese et al., 1997), respectively. The interfacial residues for EIN in the EIN–HPr complex (PDB code 3EZA; Garrett et al., 1999) are as follows: residues 67–68, 71–72, 74–76, 78–79, 82–85, 108, 111, 114–115, 118, 120, 122–123, 126, 129–130 and 189. The interfacial residues for IIA^{Glc} in the IIA^{Glc}–HPr complex (PDB code 1GGR; Wang et al., 2000) are as follows: residues 37–41, 45–46, 68–69, 71–72, 78–80, 86–88, 90, 94, 96–97, 99, 109, 141 and 144.

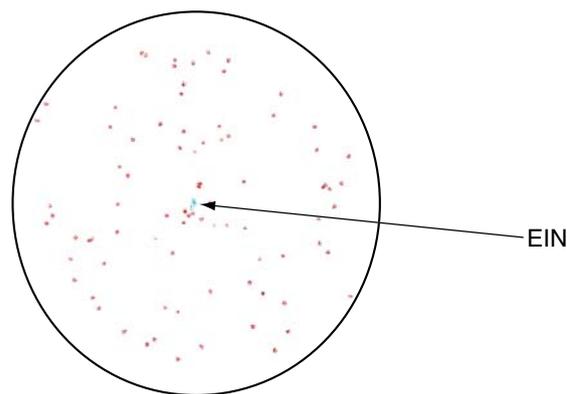


Figure 3. Starting configurations used for the EIN–HPr docking calculations. There are 100 random orientations of HPr (red) relative to EIN (cyan) in a sphere of diameter 1000 Å.

Table 3. Statistics for control calculations using highly ambiguous distance restraints (HADR) as an attractive potential between the interfaces

	EIN–HPr ^{a,b}		IIA ^{Glc} –HPr ^a	
	chain A	chain B	chain A	chain B
NOE only ^c	93/0/0	99/3/0	100/0/0	100/0/0
NOE + HADR(1) ^d	32/23/23	43/33/23	100/9/0	100/22/0
NOE + HADR(2) ^d	81/19/19	98/30/15	99/15/8	100/26/14
NOE + $R_{\text{gyr}}(\text{overall})$	86/21/21	85/28/28	99/3/3	99/8/8
NOE + $R_{\text{gyr}}(\text{overall})$ + HADR(1) ^d	34/33/25	78/8/8	88/25/16	94/46/29
NOE + $R_{\text{gyr}}(\text{overall})$ + HADR(2) ^d	26/26/17	57/28/28	76/13/11	67/7/3

^a100 starting configurations were employed with HPr randomly oriented relative to EIN and IIA^{Glc}. Calculations were carried out using both the A and B chains of EIN and IIA^{Glc} found in the asymmetric unit of the respective free crystal structures. Structures that did not converge as defined by $E_{\text{repel}} > 1000 \text{ kcal mol}^{-1} \text{ \AA}^{-4}$ or $E_{\text{NOE}} > 200 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ (values that are at least two orders of magnitude larger than for the converged structures) were excluded from further consideration. The first value reported is the total number of structures remaining from the 100 calculated structures after excluding non-converged structures based on the values of E_{repel} or E_{NOE} . The second value is the number of converged structures with a backbone r.m.s. difference within 3 Å of the respective reference structure. The third value is the number of structures in the largest converged cluster (with near identical coordinate positions) with an atomic r.m.s. difference within 3 Å of the respective reference structure. ^bThe overall R_{gyr} restraint for the EIN–HPr calculations was calculated excluding the α/β domain of EIN (residues 1–34 and 143–259). This is because EIN is a very elongated molecule and HPr only binds to the α domain (residues 35–142) (Garrett et al., 1999). If all EIN residues are used to calculate $R_{\text{gyr}}(\text{overall})$, HPr is artificially forced towards the middle of EIN, away from the binding site. ^cThe NOE restraints are those listed in Table 2. Given that all NOEs originate from a single point (the δ -methyl of an Ile residue), it is impossible to derive a unique solution based on the NOE data alone, since the corresponding distances are invariant to rotations perpendicular (as well as approximately perpendicular) to the interface plane. ^dHADR(1) and HADR(2) are highly ambiguous distance restraints making use of interfacial residues with accessible surface areas $\geq 50\%$ and $\geq 5\%$, respectively.

IIA^{Mtl}–HPr (Cornilescu et al., 2002), starting from the respective free coordinates and the calculations also resulted in a uniformly converged structure with an accuracy of $\sim 2 \text{ \AA}$ (data not shown).

Control calculations using highly ambiguous distance restraints

In earlier work (Dominguez et al., 2003; Clore and Schwieters, 2003) the attractive potential between interfaces was described by a set of highly ambiguous distance restraints (HADRs), each consisting of a $\Sigma(r^{-6})^{-1/6}$ sum of 400–3000 individual distances. The results of several control calculations using HADRs are summarized in Table 3. The protocol follows that used above except that (a) the ‘reduced’ E_{rgyr} term restricted to interfacial residues is *not* employed but instead, where indicated, a R_{gyr} restraint for the whole complex is applied with a target value calculated using $2.2 N^{0.38}$ where N is the total number of residues, and (b) the HADR restraints are included in the E_{NOE} term. Because the convergence rate is low and because the conformation that is closest to the correct structure does not necessarily have the lowest value of the target function, it was impos-

sible to reliably distinguish the correct from incorrect solutions.

Concluding remarks

In conclusion, we have demonstrated that a simple approach making use of an attractive potential in the form of a ‘reduced’ R_{gyr} restraint confined to the solvent exposed residues of the two interfaces coupled with very sparse intermolecular interproton distance restraints (as few as 3 from a single point) can be used to reliably dock protein–protein complexes at a backbone accuracy of $\sim 2 \text{ \AA}$. Reduction in the intermolecular distances below three can still lead to correct docking but the convergence rate is reduced and it can be difficult to distinguish correct from incorrect solutions. We would note, however, that if the starting coordinates employed comprise those of the complex simply pried apart and placed in random orientations relative to each other, the structures with the lowest value of the ‘reduced’ R_{gyr} potential always correspond to the correct solution even in the complete absence of any intermolecular distance restraints.

Although no attempt was made to optimize side-chain conformations owing to lack of experimental data, their optimization by conjoined rigid body/torsion angle dynamics (Schwieters and Clore, 2001) incorporating a potential of mean force for sidechain–sidechain interactions at protein interfaces (Wang et al., 2005) can be readily envisaged. Importantly, the docked structures provide a starting point for interpreting intermolecular NOEs observed in conventionally labeled samples. The approach presented here may find its application in areas other than NMR since sparse intermolecular distances can be derived from double cycle mutagenesis experiments (Schreiber and Fersht, 1995), cross-linking combined with mass spectrometry (Trester-Zedlitz et al., 2003; Schulz et al., 2004), or fluorescence energy transfer, and protein–protein interfaces can be delineated from alanine scanning mutagenesis (Jones et al., 1998; Wells, 1996).

Acknowledgments

This work was supported by the intramural program of NIH, NIDDK and the AIDS Targeted Antiviral Program of the Office of the Director of NIH (to G.M.C.).

References

- Bax, A. and Grishaev, A. (2005) *Curr. Op. Struct. Biol.*, **15**, 563–570.
- Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (2005) *Curr. Op. Chem. Biol.*, **9**, 501–508.
- Cai, M., Williams, D.C., Wang, G., Lee, B.R., Peterkofsky, A. and Clore, G.M. (2003) *J. Biol. Chem.*, **278**, 25191–25206.
- Clore, G.M. (2000) *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 9021–9025.
- Clore, G.M. and Gronenborn, A.M. (1995) *Crit. Rev. Biochem. Mol. Biol.*, **30**, 351–385.
- Clore, G.M. and Gronenborn, A.M. (1998) *Trends Biotech.*, **16**, 22–34.
- Clore, G.M. and Schwieters, C.D. (2003) *J. Am. Chem. Soc.*, **125**, 2902–2912.
- Clore, G.M., Gronenborn, A.M., Nilges, M. and Ryan, C. (1987) *Biochemistry*, **26**, 8012–8023.
- Cornilescu, G., Lee, B.R., Cornilescu, C.C., Wang, G., Peterkofsky, A. and Clore, G.M. (2002) *J. Biol. Chem.*, **277**, 42289–42298.
- Delaglio, G., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 27–93.
- Dominguez, C., Boelens, R. and Bonvin, A.M.J.J. (2003) *J. Am. Chem. Soc.*, **125**, 1731–1737.
- Fahmy, A. and Wagner, G. (2002) *J. Am. Chem. Soc.*, **124**, 1241–1250.
- Feese, M.D., Comolli, L., Meadow, N.D., Roseman, S. and Remington, S.J. (1997) *Biochemistry*, **36**, 16087–16096.
- Garrett, D.S., Powers, R., Gronenborn, A.M. and Clore, G.M. (1991) *J. Magn. Reson.*, **95**, 214–220.
- Garrett, D.S., Seok, Y.-J., Peterkofsky, A., Gronenborn, A.M. and Clore, G.M. (1999) *Nature Struct. Biol.*, **6**, 166–173.
- Goto, N.K., Gardner, K.H., Mueller, G.A., Willis, R.C. and Kay, L.E. (1996) *J. Biomol. NMR*, **13**, 369–374.
- Jia, Z., Quail, J.W., Waygood, E.B. and Delbaere, L.T. (1993) *J. Biol. Chem.*, **268**, 22490–22501.
- Jones, J.T., Ballinger, M.D., Pisacane, P.I., Lofgren, J.A., Fitzpatrick, V.D., Fairbrother, W.J., Wells, J.A. and Sliwkowski, M.X. (1998) *J. Biol. Chem.*, **273**, 1667–1674.
- Kato, K., Matsunaga, C., Igarashi, T., Kim, H., Odaka, A., Shimada, I. and Arata, Y. (1991) *Biochemistry*, **30**, 270–278.
- Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1999) *J. Am. Chem. Soc.*, **121**, 2337–2338.
- Kuszewski, J., Schwieters, C.D., Garrett, D.S., Byrd, R.A., Tjandra, N. and Clore, G.M. (2004) *J. Am. Chem. Soc.*, **126**, 6258–6273.
- Liao, D.I., Silverton, E., Seok, Y.-J., Lee, B.R., Peterkofsky, A. and Davies, D.R. (1996) *Structure*, **4**, 861–872.
- Mendez, R., Leplae, R., Lensink, M.F. and Wodak, S.J. (1995) *Proteins*, **60**, 150–169.
- Mueller, G.A., Kirby, T.W., deRose, E.F. and London, R.E. (2003) *J. Magn. Reson.*, **165**, 237–247.
- Nilges, M., Gronenborn, A.M., Brünger, A.T. and Clore, G.M. (1988) *Protein Eng.*, **2**, 27–38.
- Russell, R.B., Alber, F., Aloy, P., Davis, F.P., Korkin, D., Pichaud, M., Topf, M. and Sali, A. (2004) *Curr. Op. Struct. Biol.*, **14**, 313–324.
- Schulz, D.M., Ihling, C., Clore, G.M. and Sinz, A. (2004) *Biochemistry*, **43**, 4703–4715.
- Schwieters, C.D. and Clore, G.M. (2001) *J. Magn. Reson.*, **152**, 288–302.
- Schwieters, C.D., Kuszewski, J., Tjandra, N. and Clore, G.M. (2003) *J. Magn. Reson.*, **160**, 66–74.
- Schwieters, C.D., Kuszewski, J. and Clore, G.M. (2006) *Progr. Nucl. Magn. Reson. Spectr.*, **48**, 47–62.
- Schreiber, G. and Fersht, A.R. (1995) *J. Mol. Biol.*, **248**, 478–486.
- Suh, J.-Y., Cai, M., Williams, D.C. and Clore, G.M. (2006) *J. Biol. Chem.*, **281**, 8939–8949.
- Takeuchi, K. and Wagner, G. (2006) *Curr. Op. Struct. Biol.*, **16**, 109–117.
- Tang, C., Iwahara, J. and Clore, G.M. (2005) *J. Biomol. NMR*, **33**, 105–121.
- Trester-Zedlitz, M., Kamada, K., Burley, S.K., Fenyó, D., Chait, B.T. and Muir, T.W. (2003) *J. Am. Chem. Soc.*, **125**, 2416–2425.
- van Dijk, A.D.J., Boelens, R. and Bonvin, A.M.J.J. (2005) *FEBS J.*, **272**, 293–312.
- Wang, C., Shueler-Forman, O. and Baker, D. (2005) *Protein Sci.*, **14**, 1328–1339.
- Wang, G., Louis, J.M., Sondej, M., Seok, Y.-J., Peterkofsky, A. and Clore, G.M. (2000) *EMBO J.*, **19**, 5635–5649.
- Williams, D.C., Cai, M. and Clore, G.M. (2004) *J. Biol. Chem.*, **279**, 1449–1457.
- Williams, D.C., Cai, M., Suh, J.-Y., Peterkofsky, A. and Clore, G.M. (2005) *J. Biol. Chem.*, **280**, 20775–20784.
- Wells, J.A. (1996) *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 1–6.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
- Yu, L., Sun, C., Song, D., Shen, J., Xu, N., Gunasekera, A., Hajduk, P.J. and Olejniczak, E.T. (2005) *Biochemistry*, **44**, 15834–15841.