PEDS Advance Access published October 25, 2005

Protein Engineering, Design & Selection
doi:10.1093/protein/gzi070

Modelling the structure of latexin–carboxypeptidase A complex based on chemical cross-linking and molecular docking

Dmitri Mouradov[1], Ari Craven[1], Jade K.Forwood[1], Jack U.Flanagan[2,3], Raquel García-Castellanos[5], F.Xavier Gomis-Rüth[5], David A.Hume[1,2,3,4], Jennifer L.Martin[1,2], Bostjan Kobe[1,2] and Thomas Huber[1,6,7]

[1]School of Molecular and Microbial Sciences, [2]Institute for Molecular Bioscience, [3]Cooperative Research Centre for Chronic Inflammatory Diseases, [4]ARC Special Research Centre for Functional and Applied Genomics and [6]Advanced Computational Modelling Centre, Department of Mathematics, The University of Queensland, Brisbane, Queensland 4072, Australia and [5]Institut de Biologia Molecular de Barcelona, Centre d'Investigació i Desenvolupament, Consell Superior d'Investigacions Científiques, c/Jordi Girona 18–26, E-08034 Barcelona, Spain

[7]To whom correspondence should be addressed, at the Advanced Computational Modelling Centre, Department of Mathematics. E-mail: huber@maths.uq.edu.au

We have determined the three-dimensional structure of the protein complex between latexin and carboxypeptidase A using a combination of chemical cross-linking, mass spectrometry and molecular docking. The locations of three intermolecular cross-links were identified using mass spectrometry and these constraints were used in combination with a speed-optimised docking algorithm allowing us to evaluate more than 3×10^{11} possible conformations. While cross-links represent only limited structural constraints, the combination of only three experimental cross-links with very basic molecular docking was sufficient to determine the complex structure. The crystal structure of the complex between latexin and carboxypeptidase A4 determined recently allowed us to assess the success of this structure determination approach. Our structure was shown to be within 4 Å r.m.s. deviation of Cα atoms of the crystal structure. The study demonstrates that cross-linking in combination with mass spectrometry can lead to efficient and accurate structural modelling of protein complexes.

Keywords: chemical cross-linking/latexin–carboxypeptidaseA/mass spectrometry/molecular docking/protein complex structure

Introduction

Most cellular functions require a delicately balanced interplay of multi-protein complexes and transient protein–protein interactions. Elucidation of such protein–protein interactions at the atomic level leads us to a greater understanding of cellular processes and opens the way to regulate these processes actively, leading to many applications in biotechnology. Unfortunately, the structures of protein complexes are difficult to study with traditional structure determination methods using X-ray crystallography and NMR spectroscopy, hence there are relatively few protein–protein complexes for which the structure has been determined (Janin, 2005).

The structure of proteins in molecular complexes is often not significantly different from their structure when they are determined in isolation and they generally exhibit complementarity in shape and chemical properties at the interface (Jones and Thornton, 1996). One can therefore try to use the three-dimensional structures of the individual proteins and determine the missing information, the relative orientation of the molecules in the complex, by calculation. Molecular docking techniques have over recent decades made important methodological advances, such as employing fast Fourier (Katchalski-Katzir *et al*., 1992) or spherical harmonic (Ritchie *et al*., 1999) transforms. Useful empirical improvements that allow more reliable scoring of a large number of docking configurations (Halperin *et al*., 2002) have also been introduced. The Critical Assessment of Predicted Interactions (CAPRI) experiments (Zacharias, 2005) have been established to monitor progress and successes of molecular docking approaches; however, despite continuing incremental improvements, molecular docking remains a difficult problem and computed protein complex structures are often unreliable and of limited use.

A variety of biophysical and biochemical techniques exist that can produce rapid experimental information regarding a protein's environment and facilitate computational studies of protein–protein interactions. Molecular probe techniques, such as FRET (fluorescence resonance energy transfer) (Goedken *et al*., 2005) and EPR (electron paramagnetic resonance) (Popp *et al*., 2005) labels, are frequently used to measure selective distances between parts of a molecule. Recently, it has been shown that artificial amino acids can be selectively and efficiently incorporated into proteins for use as probes using a cell-free expression system (Ozawa *et al*., 2005). Similarly, new developments in NMR spectroscopy employ long-range electronic effects of paramagnetic ions to determine the alignment of the paramagnetic anisotropy tensor in a protein molecule (Pintacuda *et al*., 2004). By generating all possible tensor juxtapositions, it is possible to compute the relative orientation of proteins in a complex (Ubbink *et al*., 1998).

Another emerging approach to derive a set of sparse distance constraints, which can then facilitate computational structure prediction, is based on the use of chemical cross-linkers (Swaney, 1986; Friedhoff, 2005). Chemical cross-linking has been used successfully for many years to study protein interactions in virus particles (Zhu and Courtney, 1988) and other large protein complexes (Benashski and King, 2000; Rappsilber *et al*., 2000). Topological models have been derived from such cross-linking studies in the past. However, more detailed models were generally not obtained because in most cases it was not possible to determine exactly which residues had been cross-linked.

Recent advances in mass spectrometry allow the identification of the exact insertion points of low-abundance cross-links and has opened up a new perspective on the use of cross-linkers

in combination with computational structure prediction (Friedhoff, 2005). This approach is also amenable to high throughput (Young *et al*., 2000). Various groups have successfully investigated the feasibility of using chemical cross-linking as a tool for probing spatial organization of protein complexes by matching cross-links to already solved structures (Kalkhof *et al*., 2005; Tang *et al*., 2005). Other groups (Bennett *et al*., 2000; Sinz and Wang, 2001) have applied the method successfully to map out residues in the protein interaction interface. One approach to using chemical cross-linking information that does not appear to have been greatly exploited before is to combine it with molecular docking so that the cross-links are treated as explicit constraints in the calculations.

Here, we applied this strategy to characterize the mode of interaction between carboxypeptidase A (CPA) and its inhibitor latexin. Carboxypeptidases catalyse the hydrolysis of peptide bonds at the C-terminus of peptides and proteins. CPA is a metallocarboxypeptidase containing a catalytic Zn^{2+} ion and is a prototype for a family of enzymes with this activity (Vendrell *et al*., 2000). The only known mammalian carboxypeptidase inhibitor is latexin and, despite the determination of the latexin crystal structure, its mode of interaction with CPA remained unclear (Aagaard *et al*., 2005). Recently, the crystal structure of the complex between latexin and CPA4 was determined (Pallares *et al*., 2005), allowing us to assess the accuracy of the models derived from molecular docking with cross-linking restraints and to evaluate the feasibility of this cross-linking method for high-throughput structure determination of protein–protein complexes.

Materials and methods

Purification of latexin–CPA1 complex

Mouse latexin was expressed in *Escherichia coli* and purified as previously described (Aagaard *et al*., 2005). Briefly, latexin containing an N-terminal His-tag (MKHHHHHHSGA) was expressed in BL21 DE3 pLysS cells at 37°C by autoinduction (Studier, 2005) and grown until the culture reached an $OD_{600\ nm}$ of ~5. The pellet was resuspended in buffer A (50 mM phosphate buffer pH 8.0, 300 mM NaCl, 20 mM imidazole) and lysozyme was added to a final concentration of 1 mg/ml. The lysate was centrifuged at 15 000 r.p.m. (JA20 rotor) for 15 min at 4°C. The supernatant was collected and loaded on to a 5 ml Ni-NTA column, eluted using an imidazole gradient and loaded directly on to an S200 gel filtration column pre-equilibrated in gel filtration buffer (20 mM HEPES pH 7.5, 100 mM NaCl).

Bovine CPA1, purchased from Sigma (C0261), was resuspended in 15 ml of phosphate-buffered saline (PBS) and filtered through a 0.45 μm filter. The sample was then purified by gel filtration on an S200 column containing 20 mM HEPES 7.5, 100 mM NaCl and 10 μM $ZnCl_2$. Latexin and CPA1 were then combined, incubated on ice for 30 min and further purified by gel filtration using S200 gel filtration in 20 mM HEPES 7.5, 100 mM NaCl and 10 μM $ZnCl_2$. Fractions were pooled, concentrated to 30 mg/ml using a Millipore Amicon filtration device (10 000 MW cut-off) and stored at −80°C.

Cross-linking

A 100 μl volume of latexin–CPA1 complex (8 mg/ml in 100 mM HEPES, 1 M NaCl, pH 7.1) was combined with 900 μl of cross-linking solution {5 mM citrate buffer pH 5 and 2 mM BS^3 [bis(sulfosuccinimidyl) suberate] cross-linker (Sigma, S5799)} for a final pH of ~5.4 and incubated for 24 h at room temperature before the reaction was quenched using 20 μl of 20 mM Tris buffer (pH 8).

In-gel digestion and extraction

Intermolecularly cross-linked complex was purified from non-linked monomers on a Gradipore precast SDS–PAGE gel. After staining with Coomassie Brilliant Blue, the band of interest containing the cross-linked CPA1–latexin complex was excised. The band was further destained using several washes of 200 μl of 50% CH_3CN, 50 mM NH_4HCO_3. The sample was dried and incubated in 5 μl of 0.5 mg/ml trypsin (Sigma) and 200 μl of 50 mM NH_4HCO_3 at 37°C overnight.

The digested sample was centrifuged and the peptides were extracted from the supernatant with 100 μl of 60% CH_3CN–0.1% TFA, shaking at 200 r.p.m. for 30 min at 37°C. The sample was then centrifuged at 3000 r.p.m. and the supernatant pooled. The extraction process was repeated three times further. The pooled sample was dried using a SpeedVac and resuspended in 100 μl 60% CH_3CN–0.1% TFA.

Mass spectrometry

The cross-linked peptide solution was analysed using electrospray ionisation mass spectrometry (ESI-MS). The peptides were first separated by reversed-phase HPLC using a C_{18} capillary column (Agilent) and then eluted with a gradient of 0–60% (v/v) acetonitrile in 0.1% aqueous acetic acid over 45 min at a flow rate of 0.1 μl/min. The column was connected in-line to an Applied Biosystems QSTAR Pulsar mass spectrometer, which was used to record mass spectra.

Peptide assignment

The set of peaks obtained from the ESI mass spectra was analysed using an in-house program that assigns *m*/*z* values to possible cross-linked peptide fragments from amino acid sequences. Putatively assigned cross-linked fragments were then cross-checked with the original spectra for validation of real peaks by identifying multiply charged states.

Docking with distance constraints

The structures of murine latexin (1WNH) and bovine CPA1 (1M4L) were used for all docking calculations. The best docking orientation of latexin relative to CPA1 (centred at the origin) was computed by a systematic six-dimensional search over all rotations in steps of 5° and all Cartesian translations of 1.0 Å up to ±66 Å along each coordinate axis. This gives a total of $129\,168 \times 133^3$ or more than 3×10^{11} configurations. Docking calculations took ~20 min on a Pentium 4 3.0 GHz computer (1 GB RAM, 512 kB cache). Given the cross-linker reagent used here, the maximal Cα–Cα distance between cross-linked lysine residues is estimated as 25 Å; models with distances >25 Å were therefore immediately excluded from further analysis. This screening of configurations can be performed very efficiently. To save time, a pre-screen was used in which only the coordinates of those residues involved in the constraints were rotated and translated. Only when the constraints were met were the rest of the coordinates rotated and translated and a full analysis was carried out. A linear scaling grid cell algorithm with geometric hashing was used further to check for any intermolecular residue pairs in close spatial proximity and thus to exclude those models with steric overlap, defined here when

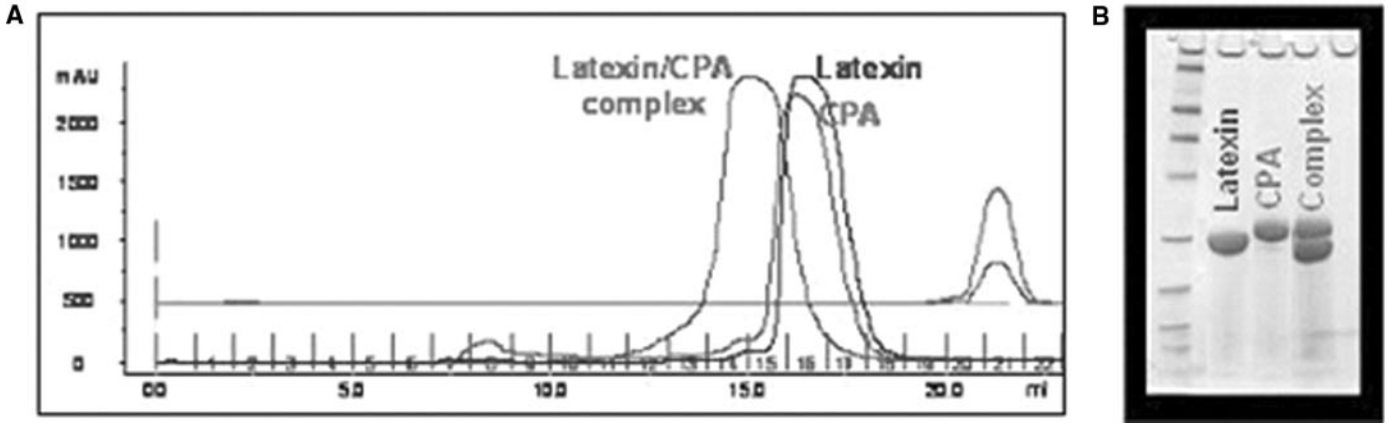


Fig. 1. Purification of CPA1, latexin and the CPA1–latexin complex. (**A**) Elution profiles of latexin, CPA1 and latexin–CPA1 complex from an S200 26/60 gel filtration column (Pharmacia). (**B**) SDS–PAGE analysis of the eluates confirming the purity of the samples. Please note that a colour version of this figure is available as Supplementary data at *PEDS* Online, where latexin is blue, CPA1 is red and latexin–CPA1 complex is brown.

Table I. Identified BS^3 cross-linked peptides from the CPA1–latexin complex[a]

Experimental mass (Da)	Predicted mass (Da)	Sequence of cross-linked peptides (latexin, CPA1)	Charged states
3709.3454	3710.06068	LFLVQTVQQAS@EDIPGR + TELNQVA@SAVAALK	+3, +4
5081.9239	5081.51584	PVQHLAWVACGYVMWQNSTEDTWY@MLK + SAVAAL@SLYGTSYK	+4, +5
4524.53	4524.20083	LFLVQTVQQAS@EDIPGR + NWDAGFG@AGASSSPCSETYHGK	+3, +4

[a]The experimental and predicted masses are monoisotopic. The @ character represents the cross-linked lysine residue. Two different charged states were identified for each cross-link.

Cα centres come closer than 3.5 Å to each other. At the beginning of the calculation, atoms of the fixed molecule are assigned into cubic cells with dimensions of the steric overlap distance (3.5 Å). A residue in the rotated molecule is then checked for steric overlap by considering only residues from the fixed molecule which are in the same cell or the 26 adjacent cells, since only those can be within the defined distance. Coordinates of the residues in each cell are stored in a linked list for which the start positions are retrieved by hash search using the 3D cell indices to construct search keys. Models were scored by a simple hydrophobic energy score that counts the number of contacts (<8 Å) between hydrophobic amino acids (A, V, L, I, F, C, M, W). For each rotation, the 10 best scoring models were retained. Final models were sorted according to their hydrophobic score and the 1000 best models considered. Models were grouped based on the root mean square deviations (RMSD) of the coordinates of Cα atoms after optimal superposition using *k*-medoids clustering (de Hoon *et al*., 2004) and the 10 best scoring models from each cluster were taken as the representative ensemble of the group. RMSD values were calculated by considering the backbone Cα atoms of both the CPA1 and latexin molecules.

Results

Purification of latexin–CPA1 complex

Size-exclusion chromatographic profiles demonstrate that latexin forms a stable complex with CPA1. Individually, the proteins elute from a gel filtration column consistent with being monomeric species. When combined in an equimolar ratio, the proteins elute as a single peak corresponding in size to the latexin–CPA1 complex (Figure 1A). SDS–PAGE analysis of the eluates confirms the presence of both proteins in the complex (Figure 1B). The sequence of latexin was confirmed by sequencing. The sequence of bovine CPA1 supplied by Sigma was not assigned unambiguously by the manufacturer; we found that the sequence was consistent with PDB code 1M4L by MALDI-TOF mass spectrometry.

Intermolecular cross-linking of latexin–CPA1

When the complex was treated with BS^3 cross-linking reagent, SDS–PAGE analysis under protein-denaturing conditions showed a strong band that corresponded to the combined masses of latexin and CPA1. Our ESI-MS analysis of peptide fragments after tryptic in-gel digestion of this band confirmed three intermolecular cross-linked peptides (Figure 2), which are summarized in Table I. Given that cleavage by trypsin is not observed after lysine residues that have been chemically modified by the cross-linking reagent, all three observed peptides resulted from fully digested peptides. The assigned peptides were confirmed by the observation of multiple charge states in the ESI mass spectra in the measured mass range of 400–2500 Da. The differences between experimentally measured and calculated *m*/*z* values were 0.715, 0.329 and 0.408 Da, with <0.02% relative error. No other peptides from either latexin, CPA1 or any common contaminates, such as keratin, were found in these mass ranges. The cross-linked peptides associated with mass of 5081.9239 Da contained two methionine residues and, owing to the formation of sulfoxides, an additional mass at +32 Da is observed with ∼8 times higher intensity. However, the exact mass of the oxidized fragment could not be determined owing to its overlap with a neighbouring mass peak (Figure 2B).

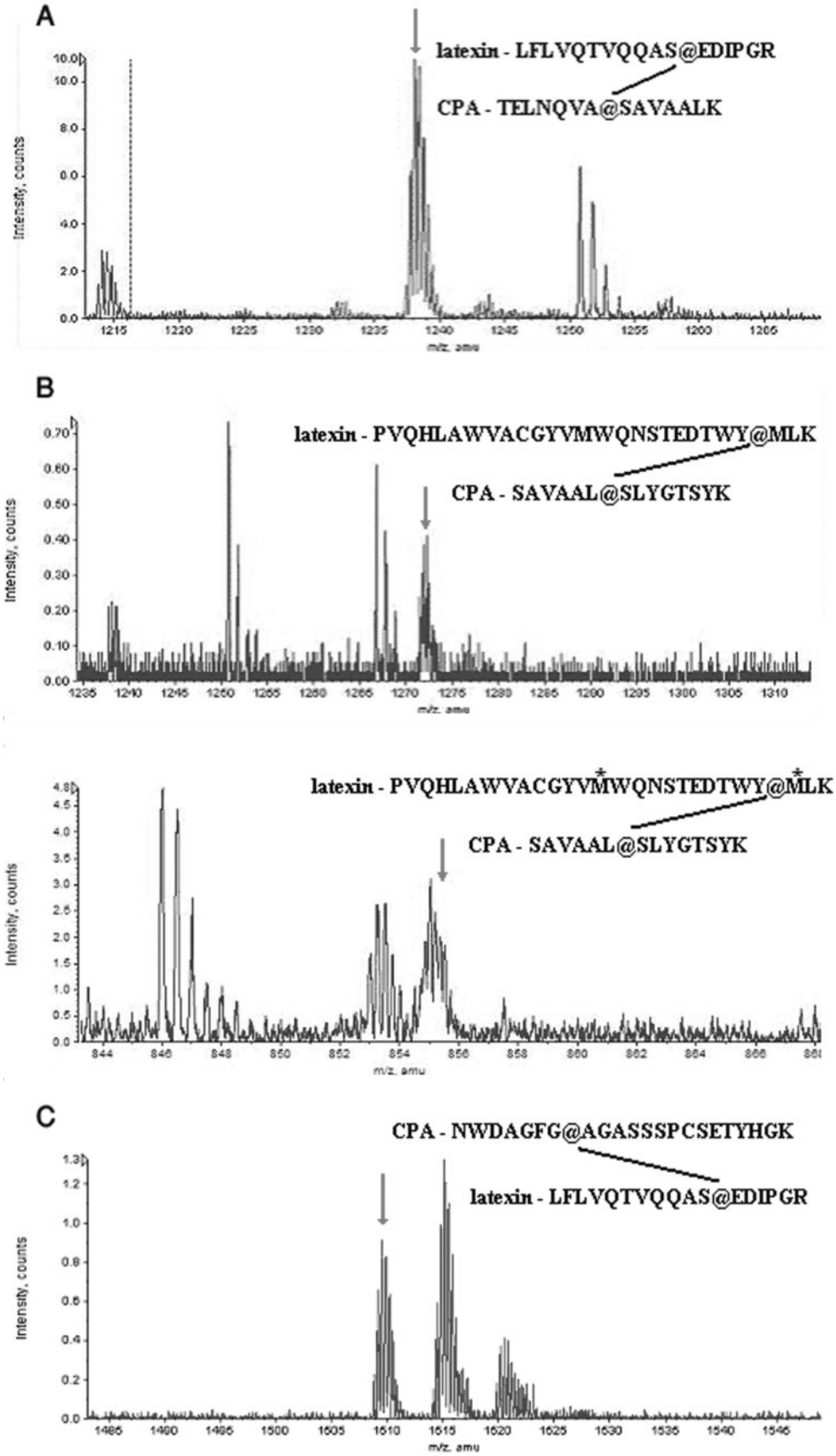
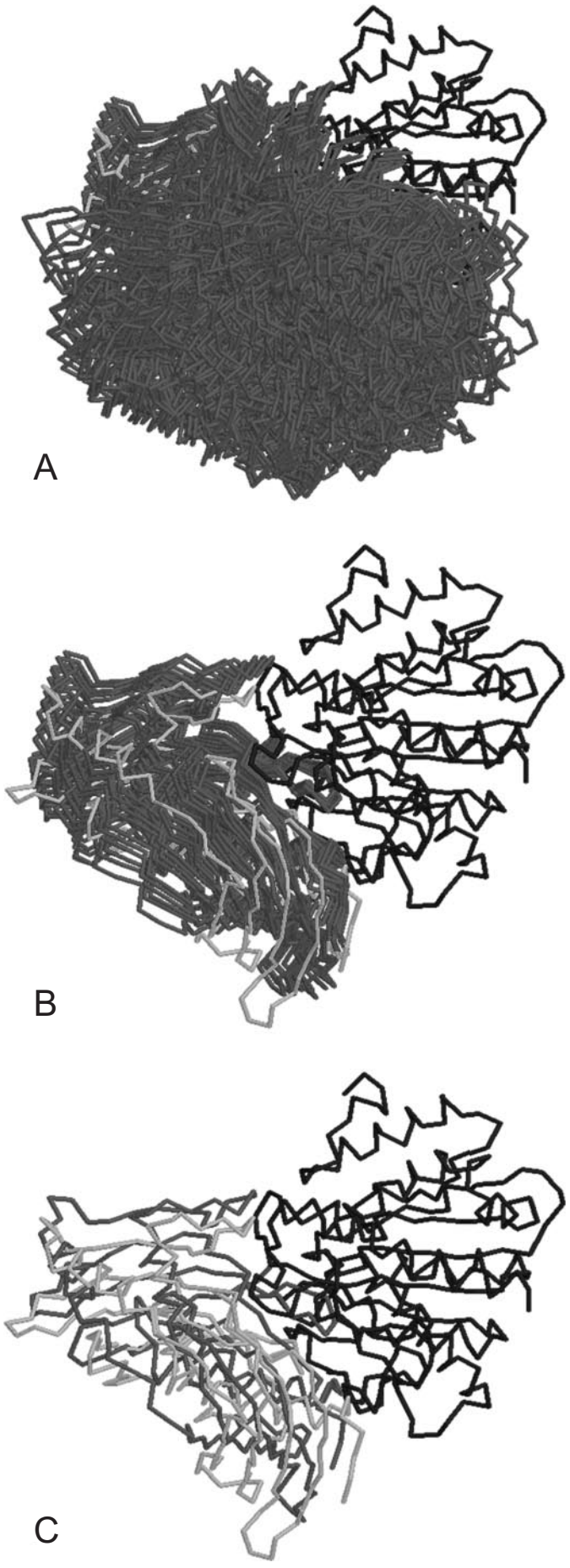


Fig. 2. ESI mass spectra of the three identified cross-linked fragments, indicated in panels (**A**)–(**C**) (the @ character indicates a cross-linked Lys). Intensity vs *m*/*z* peaks are shown (the HPLC spectra are omitted). The three peaks were taken from different chromatographic fractions. Two spectra are shown for the cross-link assigned to 5081.9239 Da (B). The top spectrum represents the unmodified fragments and the bottom spectrum shows the mass signal from two sulfoxide formations of the two methionines present in the cross-linked sequence (*). Please note that a colour version of this figure is available as Supplementary data at *PEDS* Online.



Docking of latexin–CPA1 with cross-linking constraints

Our derived cross-links imposed important distance constraints on the relative orientation of the two proteins in the complex. From the more than 3×10^{11} possible configurations, only 0.13% satisfied all three constraints and a further 99.75% of those conformations had steric overlap. The distance constraints from our cross-links are, however, not sufficient by themselves to define a single docking mode of latexin with CPA1. The final 1000 best scoring models exhibited a significant variation in structure of up to 17 Å RMSD between models and they covered large areas of putative interfaces on both molecules (Figure 3A). The models also segregate into more distinct docking modes and could be loosely grouped into 10 clusters. The population between clusters range from 19 (cluster 6) for the smallest to 350 (cluster 1) for the largest cluster. The increase in both total and hydrophobic contacts formed upon latexin–CPA1 complex formation for each cluster type is reported in Table II as an average over the ensemble of the 10 best scoring models in each cluster. The total number of contacts formed correlated directly with the size of the protein–protein interaction surface and was comparable in size for all 10 clusters. Furthermore, the number of hydrophobic contacts made by the latexin–CPA1 complex provides the basis for a simple energy-based discrimination of models. In all but one cluster, the average gain of hydrophobic contacts upon complex formation was between 8 and 10. However, models from cluster 1 formed an average of 19 hydrophobic contacts, nearly twice as many as models from any other cluster. Energetically, this cluster was clearly favoured over all other docking modes.

As the structure of the human latexin–human CPA4 complex was recently determined (Pallares *et al*., 2005), our 'low-resolution' structure could be compared directly with the high-resolution crystallographic structure. Figure 4 shows the RMSD of the Cα atoms between the crystal structure and the best 1000 models plotted against hydrophobic scores. It can be seen that the hydrophobic score discriminates between models that have been pre-screened to satisfy cross-linking and steric overlap constraints. The average RMSD between the 10 best scoring structures from cluster 1 and the crystal structure is 3.85 Å (Figure 3B) and the best scoring docked structure based on hydrophobic interactions has an RMSD of 3.74 Å when compared with the crystal structure (Figure 3C). It should be noted that this accuracy was achieved with a very simple, coarse-grained scoring function, which is based entirely on yes/no-type hydrophobic contacts between equal-sized Cα atoms and no further refinement of the models was performed. In addition, docked models with lower RSMDs were identified (Figure 4, models to the left of the best scoring structure); however, these had fewer hydrophobic interactions and hence scored lower based on our strategy.

Discussion

The central idea behind the proposed hybrid method for 'low-resolution' structure determination of protein complexes is to

Fig. 3. The crystal structure of the latexin (light grey)–CPA1 (black) complex with the orientation of the latexin molecule(s) from (**A**) top 1000 scoring docked structures (dark grey), (**B**) the top scoring cluster (dark grey) and (**C**) the top scoring docked structure (dark grey). RasWin 2.6 beta-2a was used to create these figures. Please note that a colour version of this figure is available as Supplementary data at *PEDS* Online, where latexin is green, CPA1 is blue and the top scoring docked structures and clusters are red.

use distance constraints between inter-protein residues of the complex within a molecular docking algorithm. These distance constraints are derived from cross-linking experiments, where identified cross-linked residues must be within the maximum cross-linking distance of the linker. The work reported here shows that the interaction mode of structurally characterized proteins that form a complex can be determined with a limited number of constraints when accompanied by molecular docking. Once the generic cross-linking methodology had been optimized, it took a ~5 days (3 days for cross-linker insertion and identification, 2 days for docking) to derive the interaction between the two interacting proteins, using ~0.2 mg of the protein complex.

Table II. Top 10 scoring structures in each cluster based on hydrophobic contacts: average number of contacts, hydrophobic contacts and RMSD to the crystal structure

Cluster	Total contact gain	Hydrophobic contact gain (average)	RMSD (nm)
1	143.3	19.0	0.384
2	137.0	9.9	1.514
3	97.7	9.7	1.619
4	124.6	9.5	1.487
5	118.3	9.3	1.434
6	116.2	9.3	1.325
7	99.4	9.0	1.244
8	95.2	8.1	1.501
9	70.2	8.0	1.573
10	108.8	7.8	1.233

A similar study has been undertaken investigating the calmodulin–melittin complex, where a low-resolution 3D structure was determined using chemical cross-links, mass spectrometry and docking (Schulz *et al*., 2004). However, there are a few major differences between the two approaches, mainly as a consequence of our desire to implement the process for high throughput applications. First, we chose to use one cross-linker rather than three to probe for residue distances. This means that additional distance information is lost, but the use of just one cross-linker allows a faster and simpler protocol. The second major difference is the way in which the derived distance constraints were used to generate the three-dimensional data. Schulz *et al*., used a heuristic conjoined rigid body/torsion angle simulated annealing protocol to solve the structure, whereas we applied rigid body docking by exhaustive search to generate our structure. Minimal conformational changes are often observed within the individual subunits upon complex formation between well-structured proteins (Lo Conte *et al*., 1999); in such cases, a rigid body approach is applicable. However, for interactions between small proteins (peptides) such as the calmodulin–melittin complex, significant structural changes can be expected and this flexibility must be accounted for in the docking process. This additional variable requires more constraints to determine the structure of the complex with reasonable accuracy. By contrast, only six degrees of freedom (three rotations and three translations) need to be defined in the rigid body docking approach and we have shown that this can be achieved with as few as three distance constraints. Our method can

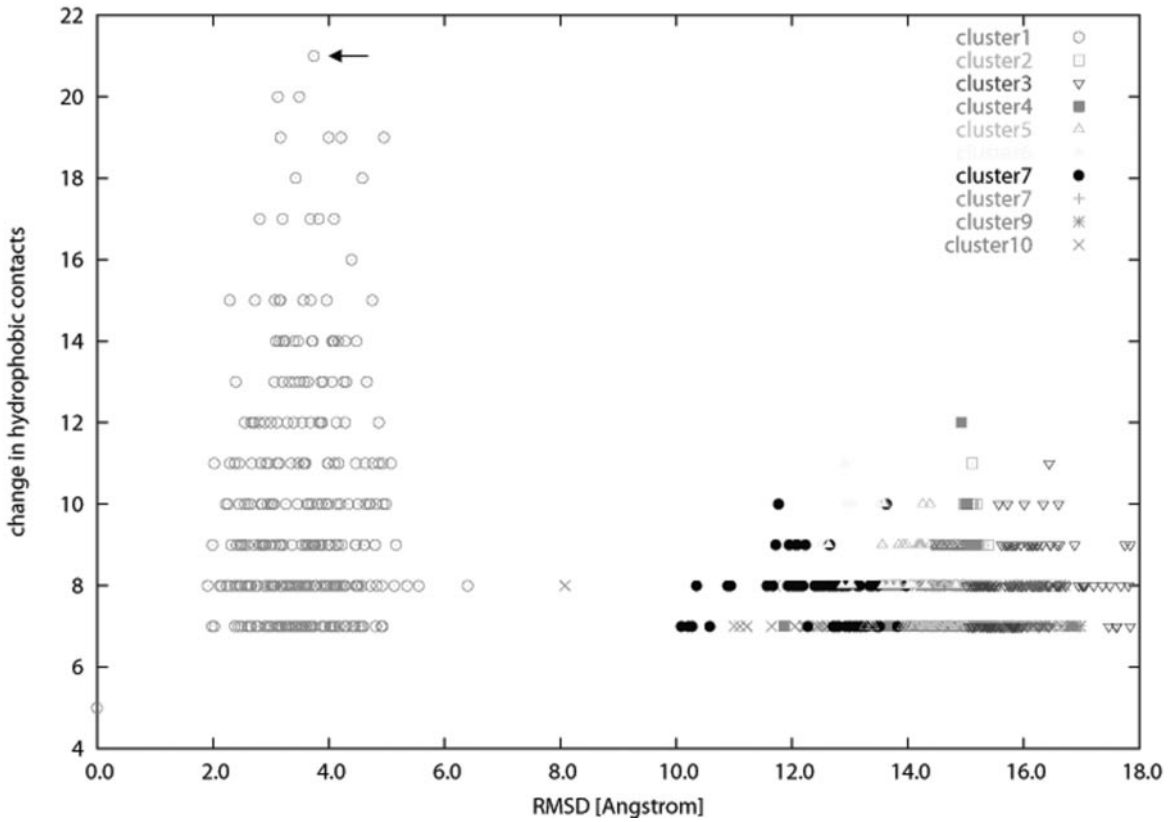


Fig. 4. RMSD of Cα atoms between the crystal structure and the best 1000 docked models plotted as a function of hydrophobic contacts gained by protein complex formation. Cluster 1 (represented by o) is clearly shown to be the best docked cluster based on RMSD when compared to the crystal structure. Our best scoring structure is highlighted with an arrow while the crystal structure is visible on the y axis. Please note that a colour version of this figure is available as Supplementary data at *PEDS* Online.

therefore be viewed as a minimalist approach to the related work of Schulz *et al*. and shows that a small number of distance constraints and a very simple yes/no scoring method can be used to deduce a reasonably accurate three-dimensional structure of the CPA1–latexin complex.

The CPA1–latexin complex proved to be a good choice for testing this methodology for several reasons. First, the structures of both components have been determined previously (Rees *et al*., 1983; Kilshtain-Vardi *et al*., 2003; Aagaard *et al*., 2005). Second, the structure of the homologous complex between human latexin and CPA4 has been solved by X-ray crystallography while our work was under way, allowing assessment of the accuracy of the method through a direct structural comparison. Third, this method relies on minimal structural change in the subunits upon the formation of the complex and the crystal structure of the human CPA4–latexin complex confirmed that there is minimal change in the conformations of the individual proteins upon complex formation.

The crystal structure (Pallares *et al*., 2005) shows that human latexin binds to CPA4 by covering the funnel-like active site of CPA4, making contacts mostly with the rim of the funnel. Latexin comprises two cystatin-like domains (Aagaard *et al*., 2005) and both domains are involved in the interaction, with the majority of contacts coming from the C-terminal domain.

The sequences of mouse and human latexin share 85% identity. After superposition, the RMSD of the mouse and human latexin structures is 0.66 Å for 217 Cα atoms [program O (Jones *et al*., 1991)]. Among the residues involved in the interaction between human latexin and CPA4 (23 residues making contacts at <4 Å), only three are not conserved in mouse latexin: Asn7 is replaced by His, Glu33 by Leu and Val161 by Leu.

The sequences of bovine CPA1 and human CPA4 share 61% identity. After superposition, the RMSD of the structure of bovine CPA1 [PDB code 1M4L (Kilshtain-Vardi *et al*., 2003)] and CPA4 from the latexin complex is 0.62 Å for 300 Cα atoms [program O (Jones *et al*., 1991)]. Out of the 25 CPA4 residues involved in the interaction between human latexin and CPA4, seven are not conserved in bovine CPA1: Val164 is replaced by Thr, Glu237 by Ser, Gln239 by Lys, Val240 by Tyr, Cys244 by Ile, Val247 by Ile and Thr276 by Arg.

There is a lack of large structural changes upon latexin–CPA1 complex formation, as inferred from the available structural data. In latexin, there are only minor differences in the main chain atoms of the segments surrounding the residues that make contacts with CPA1; for example, the main chain positions change by <1 Å in the loop connecting α3 and β6 (interacting residues 123–127) and the largest differences (just over 1 Å) are in the β8–β9 loop [interacting residues 183, 185, 187, 189–192 (Pallares *et al*., 2005)]. Some interacting side chains show different rotamer conformations in free and bound latexin, which may or may not be caused by complex formation; these include Asn125, Trp141, His185, Ile187 and Glu191. The most significant appears to be Trp141, where the rotamer conformation seen in the free mouse protein would clash with the free amino acid non-covalently bound in the CPA4 active site. The three amino acid differences between mouse and human latexin can be easily accommodated in the protein–protein interface.

The structures of free bovine CPA1 and latexin-bound human CPA4 are also remarkably similar. The largest differences are observed in the segment 237–248, partly due to amino acid changes in this area (residues 237, 239, 240, 244, 247). Thr245 shows one of the largest displacements of ∼2 Å. The largest movement occurs in Tyr248, which swings around to point in the opposite direction as it switches from an 'open' to a 'closed' or 'down' conformation due to the presence of the free amino acid in the specificity pocket of the protease (Pallares *et al*., 2005). In other interacting areas of the protein, the amino acid change of Val274 to Ile in bovine CPA1 is compensated for by a movement in the main chain, leaving the end of the side chain in a similar location in the two proteins. The Arg side chain of bovine CPA1 that replaces Thr276 in human CPA4, would clash with the valine non-covalently bound in the CPA4 active site. Finally, the side chain of Arg127 occupies a slightly different position in the two CPAs.

The best-derived structure from our docking results compares well with the solved crystal structure, having an RMSD for all Cα atoms of both proteins of less then 4 Å (Figure 3C). The orientation of latexin, in the docked structure, is rotated by several degrees compared with the crystal structure. This rotation has come around as a consequence of optimising hydrophobic interactions between the two molecules as a form of docking refinement. The N-terminus displays very little difference between the crystal structure and the docked structure; however, the rotation of the latexin forces the main the β-turn of Ala188–Ser189 in the β8–β9 loop to be placed deeper into the CPA1 active site. In addition, the α3–β6 loop (interacting residues 123–127) of latexin is shifted by ∼5 Å in the docked structure but still remains within interacting distance of the CPA1. The top-scoring docked structure also has the C-terminus of the latexin pointing in towards a pocket in the CPA1 molecule formed by residues 125–139.

The minor rotation of latexin in the docked versus crystal structure creates a total of 18 additional hydrophobic

Table III. Hydrophobic contacts of <8 Å Cα–Cα distance on both the crystal and top scoring docked structure; 18 additional hydrophobic interactions are present in the best docked structure

Docking structure hydrophobic interactions		Crystal structure hydrophobic interactions	
Latexin residue	CPA1 residue	Latexin residue	CPA1 residue
Leu125	Leu161		
Leu125	Ile182		
Leu125	Leu183		
Leu125	Leu184		
Leu125	Ile192		
Leu125	Ile193		
Trp126	Leu183		
Trp126	Ile192		
Trp126	Ile193		
Ala154	Met79		
Ile247	Ala9	Ile247	Ala9
Ile247	Ala10		
Ile247	Ala188	Ile247	Ala188
Phe279	Met160		
Phe279	Leu161		
Phe279	Leu184		
Phe279	Ile187		
Phe279	Ile192		
		Phe279	Ile192
Leu280	Leu161		
		Leu281	Phe126
Leu281	Leu161		
Ala283	Phe126	Ala283	Phe126

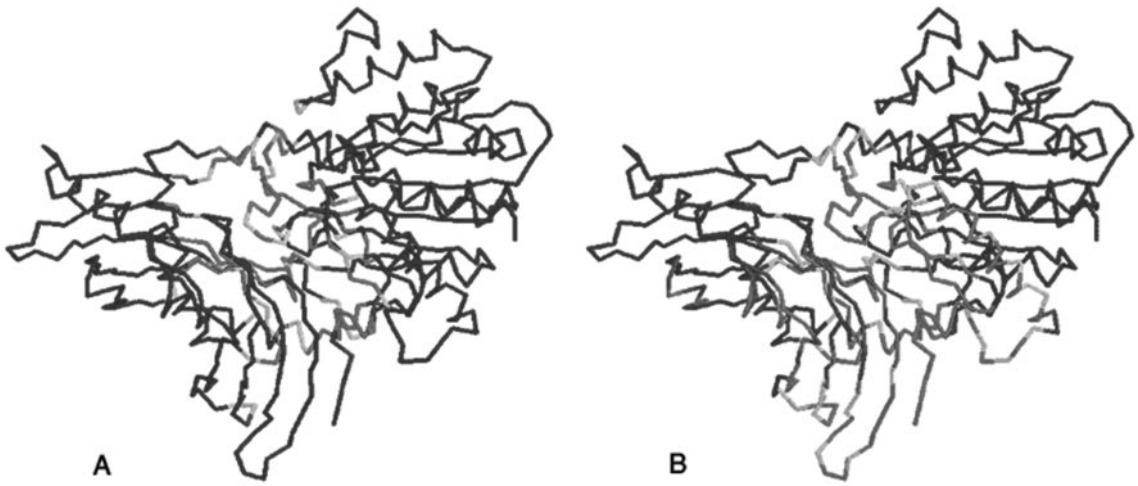


Fig. 5. The crystal structure of the complex with hydrophobic interaction hotspots from the top scoring cluster (**A**) and top 1000 docked structures (**B**). Lighter shades of grey represent hydrophobic interactions while dark grey represents no hydrophobic interactions. The model based on the top 1000 docked structures shows wide diversity of interacting sites and the model based on the top scoring cluster condenses the hotspots at the actual interacting sites. Please note that a colour version of this figure is available as Supplementary data at *PEDS* Online, where red, yellow, green and blue represent high, medium, low and no hydrophobic interactions, respectively.

interactions in the docked structure that have Cα–Cα distances of <8 Å for the two interacting proteins. Nine of the additional hydrophobic interactions involve Leu125 and Thr126 of latexin (Table III). The extra hydrophobic interactions appear to bring the two proteins slightly closer to each other than is observed in the crystal structure.

The interface of the docked structures was analysed to assess whether the accuracy was sufficient to guide the selection of putative interacting residues for mutagenesis experiments. This was performed by counting all hydrophobic interactions for each residue using the best scoring cluster and then mapping out interacting 'hot spots'. Figure 5 shows the crystal structure of the complex with these hotspots highlighted. Both interacting loops of the inhibitor identified by the crystal structure (β8–β9 and α3–β6 loops) are highlighted as hot spots in this analysis, as are the residues that form the rim of the funnel-like active site of CPA4. This analysis therefore indicates that although the RMSD with the crystal structure is ∼4 Å, the docked structure is sufficiently accurate to predict the major interaction regions of the proteins.

Conclusion

We have demonstrated the utility of 'low-resolution' structure determination of a protein–protein complex based on the structures of individual components and a limited number of chemically introduced cross-links. The structure of latexin–CPA1 was determined to 3.74 Å Cα RMSD of the crystal structure using a combination of cross-linking and mass spectrometry technology and rigid body docking with a simple scoring function. The structure can define the interface between the two molecules accurately enough to guide mutagenesis experiments probing the contribution of interacting residues and provide reagents that can be used to probe the cellular functions of the proteins. Higher resolution models can possibly be produced when models are further refined using more detailed scoring functions and all-atom models. Such approaches are currently being investigated.

References

Aagaard,A. *et al*. (2005) *Structure*, **13**, 309–317.
Benashski,S.E. and King,S.M. (2000) *Methods Enzymol.*, **22**, 365–371.
Bennett,K.L., Kussmann,M., Björk,P., Godzwon,M., Mikkelsen,M., Sørensen,P. and Roepstorff,P. (2000) *Protein Sci.*, **9**, 1503–1518.
De Hoon,M.J.L., Imoto,S., Nolan,J. and Miyano,S. (2004). *Bioinformatics*, **20**, 1453–1454.
Friedhoff,P. (2005) *Anal. Bioanal. Chem.*, **381**, 78–80.
Goedken,E.R., Kazmirski,S.L. and Bowman,G.D. (2005) *Nat. Struct. Mol. Biol.*, **12**, 183–190.
Halperin,I., Ma,B., Wolfson,H. and Nussinov,R. (2002) *Proteins*, **47**, 409–443.
Janin,J. (2005) *Protein Sci.*, **12**, 238–246.
Jones,S. and Thornton,J.M. (1996) *Proc Natl Acad. Sci. USA*, **93**, 13–20.
Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) *Acta Crystallogr. A*, **47**, 110–119.
Kalkhof,S., Ihling,C. and Mechtler,K. (2005) *Anal. Chem.*, **77**, 495–503.
Katchalski-Katzir,E., Shariv,I., Eisenstein,M., Friesman,A., Aflalo,C. and Vakser,I. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 2195–2199.
Kilshtain-Vardi,A., Glick,M., Greenblatt,H.M., Goldblum,A. and Shoham,G. (2003) *Acta Crystallogr. D*, **59**, 323–333.
Lo Conte,L., Chothia,C. and Janin,J. (1999) *J. Mol. Biol.*, **5**, 2177–2198.
Ozawa,K., Headlam, M.J., Mouradov,D., Beck,J.L., Rodgers,K.J., Dean,R.T., Huber,T., Otting,G. and Dixon,N.E. (2005) *Eur. J. Biochem.*, in press.
Pallares,I., Bonet,R., Garcia-Castellanos,R., Ventura,S., Aviles,F.X., Vendrell,J. and Gomis-Rüth,F.X. (2005) *Proc. Natl Acad. Sci. USA*, **102**, 3978–3983.
Pintacuda,G., Keniry,M.A. and Huber,T. (2004) *J. Am. Chem. Soc.*, **126**, 2963–2970.
Popp,S., Packschies,L., Radzwill,N., Vogel,K.P., Steinhoff,H.A. and Reinstein,J. (2005) *J. Mol. Biol.*, **347**, 1039–1052.
Rappsilber,J., Siniossoglou,S. and Hurt,E.C. (2000) *Anal. Chem.*, **72**, 267–275.
Rees,D.C., Lewis,M. and Lipscomb,W.N. (1983) *J. Mol. Biol.*, **168**, 367–387.
Ritchie,D.W. and Kemp,G.J.L. (1999) *J. Comput. Chem.*, **20**, 383–395
Schulz,D.M., Ihling,C., Clore,G.M. and Sinz,A. (2004) *Biochemistry*, **43**, 4703–4715.
Sinz,A. and Wang,K. (2001) *Biochemistry*, **40**, 7903–7913.
Studier,F.W. (2005) *Protein Expr. Purif.*, **41**, 207–234.
Swaney,J.B. (1986) *Methods Enzymol.*, **126**, 613–636.
Tang,X.T., Munske,G.R. and Siems,W.F. (2005) *Anal. Chem.*, **77**, 311–318.
Ubbink,M., Ejdeback,M., Karlsson,B.G. and Bendall,D.S. (1998) *Structure*, **6**, 323–335.
Vendrell,J., Querol,E. and Aviles,F.X. (2000) *Biochim. Biophys. Acta*, **7**, 284–298.
Young,M., Tang,N., Hempel,J., Oshiro,C., Taylor,E., Kuntz,I., Gibson,B. and Dollinger,G. (2000) *Proc. Natl Acad. Sci. USA*, **97**, 5802–5806.
Zacharias,M. (2005) *Proteins*, **60**, 252–256.
Zhu,Q. and Courtney,R.J. (1988) *Virology*, **167**, 377–384.

Received July 11, 2005; revised September 1, 2005; accepted September 19, 2005

Edited by Andrej Sali