Protein Science (1995), *4*:885–899. Cambridge University Press. Printed in the USA.

Monte Carlo docking with ubiquitin

MAXWELL D. CUMMINGS,[1] TREVOR N. HART,[2] AND RANDY J. READ[1,2]

[1] Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada
[2] Department of Medical Microbiology and Infectious Diseases, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

(RECEIVED December 15, 1994; ACCEPTED February 17, 1995)

Abstract

The development of general strategies for the performance of docking simulations is prerequisite to the exploitation of this powerful computational method. Comprehensive strategies can only be derived from docking experiences with a diverse array of biological systems, and we have chosen the ubiquitin/diubiquitin system as a learning tool for this process. Using our multiple-start Monte Carlo docking method, we have reconstructed the known structure of diubiquitin from its two halves as well as from two copies of the uncomplexed monomer. For both of these cases, our relatively simple potential function ranked the correct solution among the lowest energy configurations. In the experiments involving the ubiquitin monomer, various structural modifications were made to compensate for the lack of flexibility and for the lack of a covalent bond in the modeled interaction. Potentially flexible regions could be identified using available biochemical and structural information. A systematic conformational search ruled out the possibility that the required covalent bond could be formed in one family of low-energy configurations, which was distant from the observed dimer configuration. A variety of analyses was performed on the low-energy dockings obtained in the experiment involving structurally modified ubiquitin. Characterization of the size and chemical nature of the interface surfaces was a powerful adjunct to our potential function, enabling us to distinguish more accurately between correct and incorrect dockings. Calculations with the structure of tetraubiquitin indicated that the dimer configuration in this molecule is much less favorable than that observed in the diubiquitin structure, for a simple monomer–monomer pair. Based on the analysis of our results, we draw conclusions regarding some of the approximations involved in our simulations, the use of diverse chemical and biochemical information in experimental design and the analysis of docking results, as well as possible modifications to our docking protocol.

Keywords: binding; docking; drug design; Monte Carlo; prediction; simulation; ubiquitin

A variety of computer-based methods for the simulation of biomolecular docking has been reported (for reviews see Cherfils & Janin, 1993, or Kuntz et al., 1994) and this is currently an active area of research for many groups, including our own. It seems reasonable to say, however, that the development of such methods as tools for the solution of real biological problems is just beginning. The development of more accurate and robust docking algorithms requires the study of a diverse selection of biological systems, as well as critical examination of the effects of various approximations used during the simulations. The ubiquitin conjugation system provides an opportunity to study a variety of protein–protein interactions and in the present work we report the results of docking simulations, using an algorithm under development in this laboratory, with the ubiquitin/diubiquitin system.

The regulated degradation of specific proteins is one of the fundamental processes that enable cells to change rapidly from one metabolic state to another. Covalent attachment of ubiquitin polymers to protein substrates appears to be one of the major pathways by which cellular proteins are preferentially targeted for degradation in eukaryotic cells (for reviews see Hershko & Ciechanover, 1992; Hochstrasser, 1992; Jentsch, 1992; Varshavsky, 1992). Because ubiquitin is conjugated to a variety of protein substrates, it seems reasonable that this selectivity is not a function solely of ubiquitin, but derives, at least in part, from features of the enzymes involved in ubiquitin conjugation and/or features of the substrate proteins. This is supported by the existence of a large family of ubiquitin-conjugating enzymes (E2s; Rechsteiner, 1991; Hochstrasser, 1992) and the possibility of a similarly large family of ubiquitin–protein ligases (E3s; Rechsteiner, 1991). Rechsteiner (1991) has postulated that

Reprint requests to: Randy J. Read, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada; e-mail: rndy@mycroft.mmid.ualberta.ca.

ubiquitin may act as a "movable binding site," thus facilitating the interaction, or at least spatial proximity, of proteins that are not complementary to each other.

Ubiquitin is a highly conserved protein found in all eukaryotic cells. The minor sequence variations of plant and yeast ubiquitin are confined to one region of the protein, and it has been suggested that this part of the protein surface is not involved in recognition events during conjugation and/or proteolysis (Wilkinson, 1988). This region is distant from the dimer interface in diubiquitin. Chemical modification studies have indicated that several residues may be crucial to the interactions involved in activation, conjugation, or proteolysis (Wilkinson, 1988). The information available is, at best, suggestive of the relative importance or unimportance of certain regions of the ubiquitin molecule in the various intermolecular interactions that are involved in ubiquitin-dependent proteolysis. An understanding of the features involved in the interaction of ubiquitin with various enzymes of the ubiquitin conjugation pathway would help to explain some of the differences observed among the various forms of these enzymes (see, for example, Hershko & Ciechanover, 1992; Hochstrasser, 1992).

Ubiquitin activating enzymes (E1s) and E2s form thioester linkages with the carboxy-terminal glycine of ubiquitin. The catalytic cysteine has been identified as Cys 88 in the E2 UBC1 isolated from *Arabidopsis thaliana* and is located in a region that shows a relatively high degree of sequence conservation among E2s from several sources (Cook et al., 1992b). The crystal structure of this (Cook et al., 1992a) and another E2 (Cook et al., 1993) have recently been reported. Ubiquitination sites have been mapped to specific residues or regions of two degradation target proteins for which structures have been determined (Sokolik & Cohen, 1992; Hill et al., 1993). Crystal structures have also been reported for ubiquitin and Gly A76–Lys B48 isopeptide-linked diubiquitin (entries 1ubq and 1aar, respectively, in the Brookhaven Protein Data Bank [Bernstein et al., 1977]; A and B refer to the two distinct ubiquitin monomers in diubiquitin). During the course of this work, the structure of tetraubiquitin was also reported (Cook et al., 1994).

Evidence indicates that the Gly A76–Lys B48 isopeptide bond is the linkage of major importance in the ubiquitin polymers that target substrate proteins for degradation (Chau et al., 1989; Gregori et al., 1990), and this is the only linkage observed in the diubiquitin and tetraubiquitin structures. The observed twofold pseudosymmetry of the diubiquitin structure does not allow for further extension to higher polymers; however, the tetraubiquitin structure can be extended indefinitely. The flexibility of the C-terminus of the ubiquitin molecule allows a pair of covalently linked monomers access to a variety of configurations.

The biological relevance of the various polymeric states of ubiquitin is unclear. Monoubiquitination can apparently support degradation in some cases (Gregori et al., 1985; Hershko & Heller, 1985), and E2s vary in their ability to transfer ubiquitin polymers to free and ligated (to a target protein) monoubiquitin (Chen & Pickart, 1990). Diubiquitin acts as a steady-state intermediate during synthesis of higher order polymers by an E2 (Chen & Pickart, 1990). A quantitative study of the targeting efficiency of ubiquitin polymers of varying length has not been reported. One of the subunits of the proteolytic complex that degrades ubiquitinated proteins has been shown to bind ubiquitin polymers cooperatively with respect to chain length (Deveraux et al., 1994). Although it seems clear that ligation of a relatively large multiubiquitin chain to a protein can target that protein for degradation by the 26S proteasome, the functions and relative importance of the various polymeric forms of ubiquitin are currently unknown. Cook et al. (1994) state that the crystal structure of diubiquitin probably represents the predominant solution structure, and that the polymer likely undergoes a configurational "switch" to the tetraubiquitin-like configuration when a third monomer is conjugated to the growing polymer. We can find no evidence that argues against such conclusions. Although the current work is not directly concerned with clarifying these issues, our own results are consistent with the suppositions of Cook and co-workers (1994).

Because the present work is concerned with the prediction of biomolecular complexes, it is interesting to note that the observed diubiquitin structure (Cook et al., 1992a) resembles the earlier qualitative prediction of Silver et al. (1992). With the few exceptions noted above, there is little direct structural information available regarding the nature of the ubiquitin binding sites on the enzymes involved in ubiquitin conjugation, on the target proteins to which ubiquitin is conjugated, or on the proteases that recognize ubiquitinated proteins. A similar lack of information exists regarding the affinity of ubiquitin for itself or for other proteins.

We have studied the structure of diubiquitin using the 2.3-Å resolution crystal structure of diubiquitin (1aar; Cook et al., 1992a), the 2.4-Å structure of tetraubiquitin (1tbe; Cook et al., 1994), and the 1.8-Å structure of ubiquitin (1ubq; Vijay-Kumar et al., 1987). One of our long-term goals is to predict the structure of complexes involved in ubiquitin conjugation. The ubiquitin/diubiquitin system, with which we could test our ability to predict a known answer using the structures of both the complexed and uncomplexed monomers, seemed to be a logical starting point for such studies. Given that ubiquitin is known to interact specifically with numerous apparently nonhomologous enzymes, and that the affinity of one ubiquitin molecule for another is quite low (see below), we consider this system to be an especial challenge for docking methods in general. As a bonus, our initial results indicated that the ubiquitin/diubiquitin system would be very instructive for the development and evaluation of docking strategies. Despite the marked structural similarity between the ubiquitin monomer and each of the two halves of the diubiquitin structure we were unable to predict the diubiquitin structure with the unmodified ubiquitin monomer. Truncation of a flexible residue previously implicated as being crucial to one or more aspects of ubiquitin-dependent proteolysis facilitated the prediction of a dimer configuration similar to that of the experimentally observed diubiquitin molecule.

Materials and methods

Structures

The structures of ubiquitin (1ubq; Vijay-Kumar et al., 1987), diubiquitin (1aar; Cook et al., 1992a), and tetraubiquitin (1tbe; Cook et al., 1994) were from the Brookhaven Protein Data Bank (PDB; Bernstein et al., 1977).

Hardware

All calculations were performed on a Silicon Graphics R4000 Crimson or R4000PC Indy.

Software

Docking simulations were performed with the program BOXSEARCH, which is under development in this laboratory (Hart & Read, 1992). Monte Carlo minimizations were performed with a slightly modified version of BOXSEARCH. Various tools for the analysis of docking results have been developed in this laboratory. Systematic conformational searches, energy minimizations, as well as general structure manipulation and visualization were performed with DISCOVER and various modules of the INSIGHTII program (Biosym Technologies, San Diego, California). Polar hydrogen positions were optimized with the NETWORK program (Bass et al., 1992) prior to energy minimization. Some superimpositions were done according to the method of Rao and Rossmann (1973). Surface area calculations were performed with the VADAR program (under development at the University of Alberta; D.S. Wishart, pers. comm.), which incorporates the ANAREA program (Richmond, 1984). Scatter plots were prepared with the GRAPH module of the program SETOR (Evans, 1993).

Structure preparation

Water molecules were removed and hydrogens were added to the PDB structures according to the standard method in INSIGHTII at neutral pH. Any residue deletions or side-chain truncations were done at this time. Polar hydrogens were then repositioned by the program NETWORK (Bass et al., 1992), which maximizes intramolecular hydrogen bond networks (in this case intramolecular hydrogen bonds were not affected by the deletion of the waters prior to running NETWORK). The polar hydrogen positions were then further optimized by 200 cycles of steepest descents energy minimization followed by a maximum of 200 cycles of conjugate gradient energy minimization with the CVFF forcefield in DISCOVER. Minimizations were done in vacuo with a dielectric constant of 1.0, and only hydrogen atoms were allowed to move. In the case of the two halves of the diubiquitin structure, each half was treated separately so as to avoid biasing any hydrogen positions in favor of a particular docking. We consider the structure being *docked to* to be the *target* and the structure being *docked onto the target* to be the *probe*. Because our docking protocol does not allow for covalent bonds between the target and the probe, we deleted the C-terminal residue (Gly 76) from both the target and the probe in all of our docking experiments.

Reference structures

In all of the present experiments, we had a "correct" answer that we sought in our docking simulations. For the reconstruction of diubiquitin, we superimposed the two independently prepared halves of the structure onto the experimentally determined diubiquitin structure and then subjected the probe to rigid-body Monte Carlo minimization with the annealing schedule shown in Table 1. We performed one set of experiments in which we used a copy of the target as the probe. For this experiment, as well as that involving construction of diubiquitin from two ubiquitin monomers, we followed a procedure identical to that described above. The configurational space within which the docking searches took place was identical for all of the experiments reported here.

Table 1. *Minimization schedule*

Step no.	kT (kcal/mol)	No. of runs	Max. rotation (deg)	Max. translation (Å)
1	10^{-3}	500	3.0	1.0
2	10^{-4}	1,000	1.0	0.2
3	10^{-5}	1,000	0.5	0.05

Docking

Docking simulations were performed essentially as described (Hart & Read, 1992) with the annealing schedule shown in Table 2. A docking "run" with BOXSEARCH commences with the random placement of the probe within a search space that includes all or part of the target molecule. Rigid-body Monte Carlo-based simulated annealing is then performed on the probe–target configuration, according to an annealing schedule, which specifies a fixed number of Monte Carlo steps at each temperature (Table 2). Dockings that fall below a user-specified interaction energy cutoff are written to output. A typical docking experiment consists of several thousand such "runs." All of the present experiments were performed in a 49-Å cube that excluded one "face" of the target and allowed for all possible orientations of the probe relative to a large part of the target surface (see below).

Analysis—Reconstruction of diubiquitin from its two halves

All dockings were compared to the appropriate reference structure on the basis of energy and all-atom RMS differences. Cluster analysis was used to group the dockings into clusters or families. We saved the lowest energy member of each family and

Table 2. *Annealing schedule*[a]

Step no.	kT (kcal/mol)	No. of runs	Max. rotation (deg)	Max. translation (Å)
1	10	5	18	5.0
2	8.0	5	18	5.0
3	6.0	5	18	5.0
4	4.0	5	18	5.0
5	2.0	5	18	5.0
6	1.0	5	18	5.0
7	0.5	10	18	5.0
8	0.25	10	18	5.0
9	0.1	50	9	2.5
10	10^{-4}	50	9	2.5

[a] Dockings that pass the energy cutoff after this Monte Carlo run repeat step 10 four times.

counted the number of dockings that were in each family (within 2 Å RMS of the lowest energy family member).

Analysis—Construction of diubiquitin from two ubiquitin monomers

In addition to the analyses described in the previous section, we applied several more critical data filters to the results obtained in these experiments. The rotations and translations necessary to superimpose dockings onto the appropriate reference structure were determined as a complement to the more straightforward, but at times less informative, RMS differences (Shoichet & Kuntz, 1991). The rotation necessary to superimpose a docking onto the target was also determined as a measure of the pseudo-twofold symmetry of the dockings. Similar to Shoichet and Kuntz (1991), who employed mechanistic filtering to rule out incorrect dockings, we used the PROBE:75:C to TARGET:48:NZ distance to rule out certain configurations, based on the presumed difficulty of forming the necessary isopeptide bond between distant atoms. In one case, a systematic conformational search was carried out on the Lys 48 side chain of the target as well as the flexible C-terminus of the probe. Changes in exposed surface area upon complex formation were calculated for some experiments, and we also calculated a simple energy correction based on these changes (Eisenberg & McLachlan, 1986; Eisenberg et al., 1989). Changes in exposure of the various types of surface area due to complex formation were multiplied by atomic solvation parameters as described by Eisenberg et al. (1989), summed, and then added to our original interaction energies. The sole S atom in ubiquitin was treated as a polar N/O-type atom. This simple correction applies an energetic penalty for burial of polar or charged surfaces and an energetic reward for the burial of hydrophobic surfaces.

Results and discussion

Relevant biochemical information

Ubiquitin is a highly conserved protein—the sequences of all animal ubiquitins are identical, and yeast and plant ubiquitin each have three conservative substitutions (giving a total of four variant sites—19, 24, 28, 57; Ozkaynak et al., 1984; Vierstra et al., 1986). Yeast ubiquitin is fully active in assays of ubiquitin activation as well as ubiquitin-dependent proteolysis in animal-derived in vitro systems (Wilkinson et al., 1986). Oat ubiquitin is active in ubiquitin activation but stimulation of protein degradation has not been reported. It is expected to be fully active in this assay as well (Wilkinson, 1988). Wilkinson (1988) originally noted that the four variant residues of oat and yeast ubiquitin are clustered on one face of the protein and that this face, directly opposite to that of the carboxy-terminus, is probably not involved in intermolecular interactions in the ubiquitin-dependent proteolysis pathway. Subsequently, the crystal structure of diubiquitin revealed that all of these residues were distant from the dimer interface (Cook et al., 1992a). In the recently reported tetraubiquitin structure one of these variant residues (Glu 24) accepts two inter-monomer hydrogen bonds and another (Ala 28) is near an inter-monomer interface (Cook et al., 1994).

Wilkinson and co-workers have studied the effects of various chemical modifications of ubiquitin on ubiquitin activity in assays relevant to the ubiquitin-dependent proteolysis pathway (Wilkinson, 1988). Similarly, Ecker and co-workers (1987) studied the effects of various mutations on the activity of ubiquitin in in vitro protein degradation. Although these studies do not provide direct evidence of the involvement of any specific residues or regions of the protein in a particular intermolecular interaction, they do hint at the relative importance or unimportance of certain residues in such interactions. We can use such suggestions as indicators of which side chains might be involved in a protein-protein interaction at some point in the ubiquitin conjugation pathway. From their results, we concluded that we should critically examine (see following section) residues Arg 42, 72, and 74, Tyr 59, and His 68. Unfortunately, there was no such information available to us regarding mutants that could not be catalytically dimerized by ubiquitin conjugating enzyme.

From this variety of chemical and biochemical information (see above), we were able to construct a search space around the ubiquitin molecule that excluded the "variant face" of the docking target while at the same time allowing relatively unrestricted access of all possible orientations of the docking probe to a large part of the target surface. This accessible surface included all of the potentially critical residues described above. This search space excluded the possibility of obtaining dockings similar to the monomer-monomer configuration observed in the tetraubiquitin structure. However, the results of control experiments, as well as those of several analyses, suggest that the monomer-monomer configuration observed in tetraubiquitin is unlikely to be observed in a diubiquitin molecule (see below).

Relevant structural information

An obvious problem that can occur when using uncomplexed molecules to generate a complex during a rigid-body docking simulation is the clash of atoms which, in reality, could be avoided by very slight conformational adjustments (as observed, for example, in the diubiquitin structure—see below). The Lennard-Jones 6,12 potential used in our energy calculations ascribes prohibitive energy penalties to even slight atomic overlaps. In the context of the current work this means that, although the attraction due to any one side chain in a large protein-protein interface can, in many cases, be omitted without significantly altering the dockings obtained, the repulsiveness of one unfavorably positioned side chain can have a profound influence.

The flexibility of the C-terminal region of the ubiquitin molecule, described in Figure 1A, as well as the partial occupancy of these four residues, was originally noted by Vijay-Kumar and co-workers (1987). For our rigid-body docking studies, this is particularly challenging because this region of the molecule is critical to the interaction of ubiquitin with molecules to which it becomes covalently attached. The combined backbone and side-chain flexibility in this region of the ubiquitin molecule allows for a prohibitive number of accessible conformational states. We did not attempt to model this flexibility directly in the docking simulations. Instead, we deleted Gly 76 prior to performing docking, in order to eliminate the possibility of a van der Waals clash between PROBE:76:C and TARGET:48:NZ during the docking simulation. This did not result in any major configurational changes to the complex upon rigid-body Monte Carlo minimization (probe 1 in Table 3).

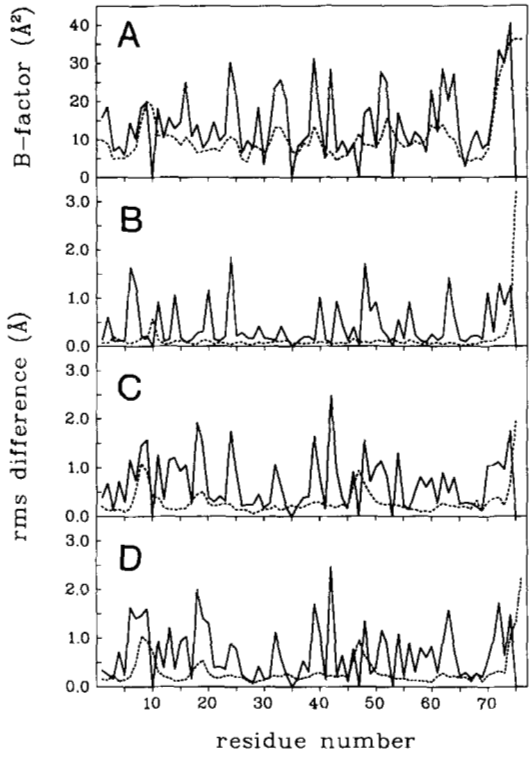

Fig. 1. Indications of flexibility in ubiquitin and the two halves of diubiquitin. Solid lines, side-chain atoms; broken lines, backbone atoms. All superpositions were between the backbone atoms of residues 1–72. Backbone values for residue 76 (omitted from plots) are 5.5 (B) and 4.8 (C). **A:** Average *B*-factors for the side-chain and backbone atoms of ubiquitin (1ubq). RMS differences between the side chains and backbones of the target and probe halves of **(B)** diubiquitin, **(C)** ubiquitin and the target, and **(D)** ubiquitin and the probe.

Flexible side chains of residues 1–72 include Glu 16, 24, and 64, Asn 25 and 60, Lys 33, Asp 39 and 52, Arg 42 and 72, and Gln 62 (Fig. 1A). Several of these residues lie on the "variant face" of the target molecule, which was excluded from our docking search (see above). These excluded residues include Glu 16 and 64, Asn 25 and 60, Lys 33, and Gln 62. Access to Glu 24 and 51 and Asp 52 was somewhat restricted. The only two relatively flexible side chains (Fig. 1A) in ubiquitin that were freely accessible to the probe molecule in our docking experiments were Asp 39 and Arg 42. In light of the difficulties we encountered when docking native (mono)ubiquitin (see below), it is interesting that of all of the relatively flexible side chains in ubiquitin the difference between side chain and backbone flexibility is greatest for Arg 42 (Fig. 1A).

Docking—Summary of experimental constraints

Prior to considering the information to be gained from the diubiquitin structure, we summarize the salient biochemical and structural information and our application of it to the design of our docking simulations as follows. First, the variant residues of plant and yeast ubiquitin suggest that we can exclude this face of the ubiquitin molecule from our search. When we constructed a search cube that excluded this face of the protein, we also excluded many of the flexible side chains in residues 1–72. This dramatically reduced the computational expense of our docking search and also eliminated many of the possible modifications that we might have considered (e.g., multiple conformations, side-chain truncations). Second, chemical and structural information indicated that flexibility in Asp 39, Arg 42 and 72, as well as residues 73–76 might create difficulties in our docking experiments. Chemical and biochemical information had previously implicated several of these residues, as well as Tyr 59 and His 68, as being potentially critical in one or more protein-protein interactions involved in the ubiquitin-dependent protein degradation pathway. The search space we constructed, which excluded the "variant face" of the docking target, allowed relatively unhindered access of all possible orientations of the docking probe to the target surface comprising all of these critical residues. Third, in a wide variety of homodimeric proteins, the majority are found to exhibit twofold symmetry (Miller, 1989). Our docking results were easily filtered to look for pseudo-twofold symmetric configurations. Fourth, we examined the nature (nonpolar, polar, charged) of the interface surfaces in our

Table 3. *Reference structures for docking experiments*[a]

Complex	$E_{complex}$[b] (kcal/mol)	E_{min}[b] (kcal/mol)	RMS[c] (Å)	Distance between centers[d] (Å²)	Rotation angle[d] (deg)	Translation along screw axis[e] (Å)	Decrease in ASA[f] (Å)		
							Nonpolar	Polar	Charged
Ub2	–	–	–	22.2	180.0	0.0	1,033 (0.68)	240 (0.16)	243 (0.16)
Ub2/probe1	−75.7	−79.2	0.7	21.9	179.0	0.2	994 (0.67)	219 (0.15)	269 (0.18)
Ub2/probe2	−46.4	−68.6	0.5	22.1	179.3	0.1	878 (0.66)	216 (0.16)	242 (0.18)
Mutant	+46.1	−39.0	2.0	23.7	176.6	1.8	641 (0.74)	158 (0.18)	72 (0.08)

[a] Data for the native diubiquitin complex (Ub2), two different diubiquitin complexes (probes 1 and 2), as well as our modified dimer (mutant).
[b] Energies before and after minimization.
[c] Movement caused by minimization.
[d] Distances between the probe and target centers were measured after minimization, as were the rotations necessary to superimpose the probe onto the target.
[e] The minimized probes were superimposed onto their respective targets by rotation about an approximate twofold screw axis. This number represents the component of the translation parallel to the screw axis.
[f] Changes in accessible surface area (ASA) were the differences between the complex and its two halves (fractions of the interface shown in parentheses).

dockings and compared these to the dimer interfaces previously characterized by other workers (Janin et al., 1988; Miller, 1989). Fifth, work by Chau and co-workers (1989) as well as others (Gregori et al., 1990) has indicated that the most important ubiquitin–ubiquitin covalent linkage occurs between Lys 48 NZ of one monomer (the target in our experiments) and Gly 76 C of the second monomer (the probe). Again our docking results were easily filtered to look for dockings that would accommodate this constraint.

Of course, the crystallographically observed structure of diubiquitin was available to us throughout the course of these docking experiments and was in fact used to aid in construction of our reference complexes. However, in designing our docking experiments, we attempted, as much as possible (see above), to use strategies that could have been deduced from previously available information, excluding the structure of diubiquitin itself. The information we used in our experimental design included the variant residues of ubiquitin, the in vitro effects of various chemical modifications to ubiquitin, as well as the dimensions of the ubiquitin monomer. Other information such as a covalent bond distance constraint and symmetry and surface considerations, as well as interaction energies and RMS differences between dockings and the appropriate reference structure, were used in the analysis of our docking results.

Docking – Criteria of success

We consider an experiment to have been successful if the appropriate reference structure is generated during the docking search *and* that structure is ranked as the lowest energy docking by BOXSEARCH. Furthermore, we would like to see that the correct answer is a popular one – that is, if we group the dockings into clusters based on RMS differences, the cluster containing the correct answer should be among the most heavily populated clusters. Because BOXSEARCH has been designed to generate all possible starting configurations with equal probability (Hart & Read, 1992), multiple visits to the more energetically favorable minima imply that our search has been reasonably exhaustive. We consider structures to be the same if the RMS difference between all atoms does not exceed 2 Å. Although our method allows for bias to be introduced prior to running the simulation, by modifying the molecules as well as by limiting the search space, once invoked, the main docking algorithm itself is completely random and free of further bias (discussed in Hart & Read, 1992).

Docking – Reconstruction of diubiquitin from its two halves

The crystal structure of diubiquitin shows that the two ubiquitin monomers in this dimer are linked by an isopeptide bond between Gly A76 C and Lys B48 NZ (Cook et al., 1992a). Ubiquitin polymers consisting solely of Gly 76–Lys 48 isopeptide bond-linked monomers have been shown to be fully competent mediators of ubiquitin-dependent proteolysis (Chau et al., 1989). Although this is the most commonly observed linkage in various systems, it is not the only one (Hochstrasser, 1992). The functions and relative importance of the various monomer–monomer linkages possible in ubiquitin polymers have yet to be determined. All of our docking simulations were aimed at generating dimers that might be covalently linked by an isopeptide bond between PROBE:76:C and TARGET:48:NZ.

We are unaware of any precise measurements of the affinity of the ubiquitin monomer for itself. If monoubiquitin self-associates in the absence of a conjugating enzyme we estimate a lower limit of 10 mM for the dissociation constant for noncovalent dimerization (calculation based on a personal communication from M. Ellison). Obviously, the affinity of monoubiquitin for itself is low, at least when the monomers are not covalently linked. The two halves of diubiquitin are linked by a flexible chain that is potentially 20 Å in length when fully extended. This linkage allows for a variety of possible monomer–monomer interactions, which combined encompass a relatively large configurational space. Conversely, Cook et al. (1992a) previously noted that it was possible to imagine a diubiquitin molecule in which the sole intermonomer interaction was the covalent bond linking the two monomers.

Our first set of experiments dealt with the two halves of the crystallographically observed diubiquitin structure. The only structural modification made in this case was the deletion of Gly 76 from both the target and the probe (probe 1). The interaction energy for this modified complex in the native configuration was calculated to be −75.7 kcal/mol (Table 3). Rigid-body Monte Carlo energy minimization of this complex led to a 0.7-Å RMS shift and a slight decrease in the calculated interaction energy (Table 3). Initial experiments showed that the RMS differences between the dockings and our reference structures were significantly decreased when the reference configuration was minimized (results not shown).

When we ran an experiment with 5,000 separate starts and an energy cutoff of −30.0 kcal/mol, 39 dockings fell below the energy cutoff (Table 4). Three of these dockings were correct with RMS differences from the reference structure of 0.6, 0.8, and 1.8 Å, and interaction energies of −74.5, −71.4, and −57.0 kcal/mol, respectively. The same experiment with 40,000 starts produced 230 dockings below the energy cutoff (Table 4) and 21 of these were within 2 Å RMS of the reference structure. These 230 dockings were divided into 131 clusters (Fig. 2), of which 86 had a single member. Twenty-eight of the 230 dockings fell into the three lowest energy clusters, and all of the dockings that were within 2 Å RMS of the reference structure fell into the two lowest energy clusters (Table 5). Various statistics for this pair of experiments are shown in Tables 4 and 5. Figure 2

Table 4. *Docking statistics for probe 1*[a]

No. of starts	No. correct[b]	RMS[c] (Å)	Energy[d] (kcal/mol)	Cluster[e]
5,000	3/39	0.6	−74.5	1/33
40,000	21/230	0.4	−76.6	1/131

[a] The reference structure had an interaction energy of −79.2 kcal/mol.

[b] Number correct/total number of dockings that passed the energy cutoff.

[c] RMS difference between the reference structure and the lowest energy docking.

[d] Interaction energy of the lowest energy docking in that experiment.

[e] Ranking of cluster containing best-fit answer/total number of clusters in this experiment.

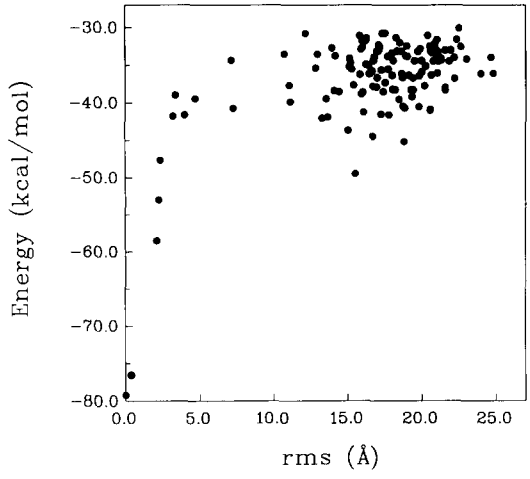

Fig. 2. Reconstruction of diubiquitin from its two halves, judged by RMS deviation from a reference structure. Forty thousand separate docking starts with probe 1 gave rise to 230 dockings with interaction energies below −30.0 kcal/mol. These were separated into 131 clusters with the three lowest energy clusters containing a total of 28 dockings. The lowest energy member of each of the 131 clusters is shown in this figure. The point with an RMS value of 0.0 represents the minimized reference configuration.

Table 5. *Top 10 clusters for large experiment with probe 1*[a]

Cluster	Energy[b] (kcal/mol)	RMS[c] (Å)	Members[d]
1	−76.6	0.4	21
2	−58.5	2.1	5
3	−53.0	2.2	2
4	−49.5	15.5	5
5	−47.6	2.4	2
6	−45.2	18.8	5
7	−44.4	16.7	2
8	−43.6	15.1	4
9	−42.0	13.3	7
10	−41.9	13.6	5

[a] Statistics listed are for the lowest energy docking in each cluster. The total number of dockings that passed the energy cutoff (−30.0 kcal/mol) in this experiment was 230.
[b] Interaction energy of the lowest energy docking.
[c] RMS deviation of the lowest energy docking from the reference structure.
[d] Number of dockings in that cluster.

clearly shows that the dockings that are distant from the reference structure are energetically unfavorable relative to the correct dockings. Figure 3 compares the orientation of the lowest energy docking obtained in this experiment to that of the reference structure.

According to all of the criteria outlined above, this series of experiments was clearly successful. Not only is our docking protocol capable of generating dimer configurations within 2 Å RMS of the crystallographically observed structure, it also ranks these dockings as the most favorable of all the dockings generated. Finally, cluster analysis of the dockings indicates that the correct docking is obtained relatively frequently. Indeed, in the large experiment described above, the low energy cluster has three times the number of members of the next most heavily populated cluster. It is particularly exciting to note that, in achieving this success, we have not had to use all of the biochemical information available to us. Specifically, we have not filtered the results to remove dockings in which the C-terminus of the probe is distant from Lys 48 NZ of the target.

Our next set of experiments was a slightly more rigorous test of our docking protocol. Instead of docking together the two halves of diubiquitin (which may have undergone minor conformational changes to become more complementary), we used a copy of the target as the probe (probe 2). For a dimer that exhibits twofold pseudosymmetry, such as diubiquitin, we expect

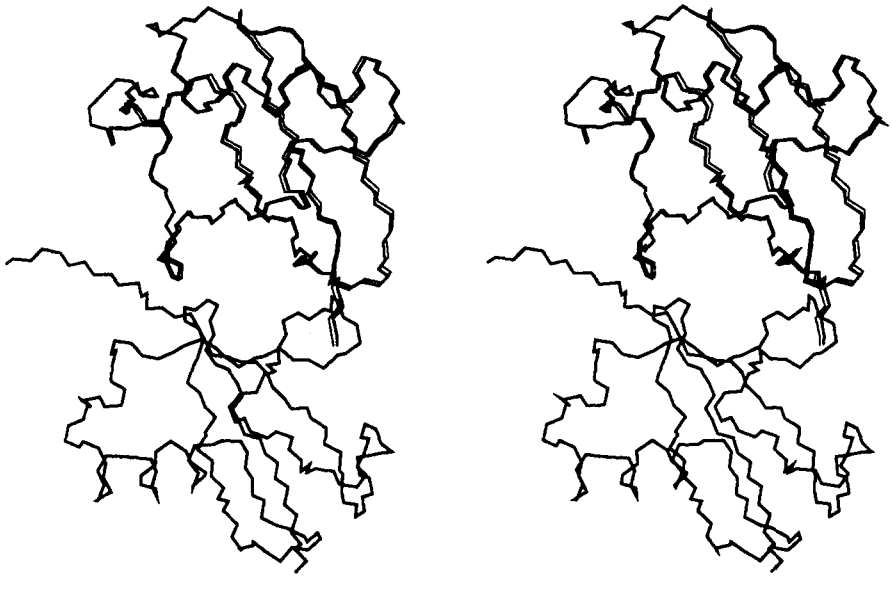

Fig. 3. Superposition of the lowest energy docking (thin line) and the reference structure (thick line) for probe 1. The interaction energy of this docking was −76.6 kcal/mol and that of the reference was −79.2 kcal/mol. Only N, C, and CA atoms are shown.

the interface regions of the two monomers to be quite similar to each other due to their similar environments. Because the dimer in this case is not perfectly symmetric, however, some differences between the two halves do exist. In general we would not expect the differences between the two halves of such a complex to be as great as the differences between the isolated monomer and either of the two halves of the complex. Figure 1B, C, and D shows that, with the exception of the C-terminus, the conformational differences between the two halves of diubiquitin are not as great as those observed for the ubiquitin monomer and the two halves of diubiquitin. This set of experiments gave us some insight into the dependence of the success of our first set of experiments on a particular set of side-chain and C-terminus conformations.

Despite the apparent decrease in favorability of this complex (probe 2 in Table 3), the results we obtained in this set of experiments (results not shown) were similar to those obtained in the first set of experiments. Using this "modified probe" we were again able to generate and correctly rank the dockings that resembled the crystallographically observed diubiquitin structure. The lowest energy cluster in the larger experiment was the most heavily populated cluster and was represented by a docking that was within 3.1 kcal/mol and 0.5 Å RMS of the reference structure. The next best cluster was 14 kcal/mol less favorable and was 17.4 Å RMS away from the reference configuration. Because the backbone conformations of the two halves of diubiquitin are virtually identical, with the exception of the C-terminus, this set of experiments showed that large modifications of the flexible parts of the probe (see Fig. 1B) did not prevent our docking protocol from generating and correctly scoring the experimentally observed diubiquitin structure. Also, once again we did not have to apply additional biochemical information during our analysis to achieve this success.

Docking – Construction of diubiquitin with two copies of (mono)ubiquitin

A more rigorous and realistic test of a docking protocol is to attempt to reproduce the experimentally determined structure of a complex using the structures of the uncomplexed (i.e., native) components of the complex. This has been achieved in a number of cases (see, for example, Goodsell & Olson, 1990; Shoichet & Kuntz, 1991; Bacon & Moult, 1992; Hart & Read, 1992). Because most computational docking protocols, including our own, allow for little or no conformational flexibility in the interacting molecules, it is not surprising that docking results obtained with uncomplexed molecules are generally not as good as those obtained with the complex components (see, for example, Shoichet & Kuntz, 1991; Bacon & Moult, 1992; Hart & Read, 1992). In the absence of structural information of some sort, the consideration of major (backbone) conformational changes that may be necessary for, or induced by, complex formation is problematic. This is especially true for the prediction of protein-protein complexes, which may involve large interfaces and dozens of flexible side chains. In the current study, we chose to deal with the flexibility of certain critical residues in two ways – by systematically searching the accessible conformational states and, similar to Shoichet and Kuntz (1991), by truncating relatively disordered residues.

A summary of potentially relevant biochemical and structural information (see above) had indicated that the flexibility (or positioning) of Asp 39, Arg 42 and 72, as well as that of the C-terminal residues 73–76, might be crucial to this simulation. Our experiments with the different probes suggested that correct docking was not dependent on a *particular* C-terminus conformation (see above). In contrast to the comparison of the two halves of diubiquitin (Fig. 1B), Figures 1C and 1D show that the conformations of Asp 39 and Arg 42 differ greatly between ubiquitin and the two halves of diubiquitin. Figure 1A shows that the side chains of these two residues are among the most flexible in ubiquitin. Arginine residues are often among the most variable and uncertain in conformation. In principle, then, the potential importance of the conformations of these residues to the success of our docking simulations could be identified from any one of several lines of evidence.

Visual inspection of the dimer revealed that, in both the target and the probe, Arg 42 is located in the middle of the dimer interface. Figure 4 shows that, although in the diubiquitin structure Arg 42 of the two halves easily accommodates dimer formation, the conformation of this side chain in the monoubiquitin structure prohibits the formation of a diubiquitin-like dimer from two copies of monoubiquitin. A docking simulation with two copies of native ubiquitin, which was in all other respects identical to our previous experiments, confirmed this (results not shown).

Modeling of side-chain flexibility during a docking simulation increases the difficulty of an already challenging problem and we did not wish to address this related issue in the current work. Instead, we took the very simplified approach of approximating the flexibility of the Arg 42 side chains by truncating them down to Ala residues. Perhaps surprisingly, this worked.

Two copies of this modified ubiquitin molecule (Arg 42 → Ala 42; Gly 76 deleted) were superimposed onto the diubiquitin structure to generate a reference dimer configuration. The interaction energy of this unminimized configuration was +46.1 kcal/mol (Table 3). Rigid-body Monte Carlo minimization gave a dimer configuration with a reduced interaction energy (Table 3). This value was still somewhat higher than those observed in our earlier experiments (Table 3). Although the RMS difference between this minimized probe and the unminimized probe was relatively large when compared to the values obtained in our earlier experiments (2.0 Å versus 0.7 or 0.5 Å; Table 3), several other statistics indicated that it represented a dimer configuration that was similar to the reference complexes we had used in our earlier experiments (Tables 3, 6). Because this experiment involved the uncomplexed monomer, the less favorable values we observed were not surprising. With the exception of the Arg 42 → Ala 42 modification, the protocol of this docking simulation was identical to that employed in our previous experiments.

In the experiment with the modified ubiquitin monomer, 40,000 starts yielded 184 dockings with interaction energies below −30.0 kcal/mol. These divided into 110 clusters, 73 of which contained a single member (Fig. 5). Figure 5 shows that two of the three lowest energy clusters are within 4.6 Å RMS of the reference structure, whereas the other low energy cluster is 23.1 Å RMS away. The lowest energy dockings of clusters 1 and 2 are shown in Figure 6. We see that these two dockings utilize radically different interfaces; these differences are further detailed by the measurements presented in Table 6 (discussed below). Fourteen of the 184 dockings are contained in the three lowest energy clusters, and cluster 1 is by far the most heavily popu-

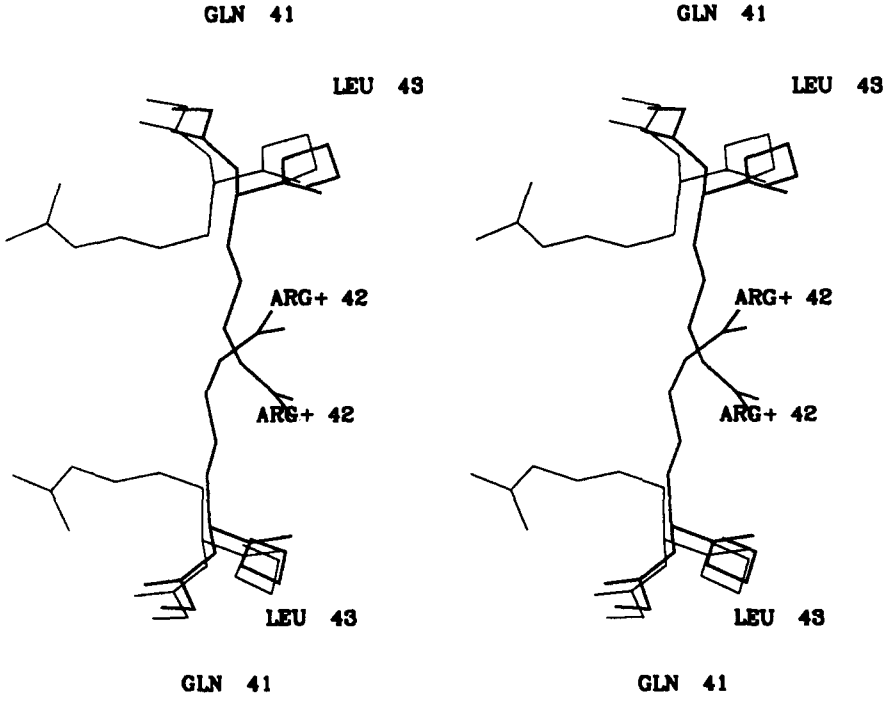

Fig. 4. Superposition of two copies of monoubiquitin (thick lines) and the two halves of diubiquitin (thin lines). The different conformation of Arg 42, which is central to the diubiquitin interface, in monoubiquitin prevented dimer formation in our initial docking simulations with two copies of monoubiquitin. The backbones of residues 41–43 and the Arg 42 side chains are shown.

lated cluster in this experiment (Table 6). Figure 5 also shows that many dockings, both near to and distant from the reference structure, have lower interaction energies than the reference structure. In contrast to our experiments involving the two halves of diubiquitin, as well as those involving two copies of one half of diubiquitin, this docking experiment was not unambiguously successful when judged solely on the basis of docking energies and RMS differences. None of the top 10 clusters

Table 6. *Top 10 clusters for experiment with modified ubiquitin*[a]

Cluster	Energy[b] (kcal/mol)	Adjusted energy[c] (kcal/mol)	RMS[d] (Å)	Members[e]	Pseudo bond distance[f] (Å)	Rotation onto target[g] Angle (deg)	Rotation onto target[g] Distance (Å)	Translation along screw axis[h] (Å)	Rotation onto probe[i] Angle (deg)	Rotation onto probe[i] Distance (Å)	Decrease in ASA[j] (Å^2) Nonpolar	Polar	Charged
Ub2/probe1	−79.2	−85.0	–	–	4.3	180.0	22.2	0.2	–	–	1,033 (0.68)	240 (0.16)	243 (0.16)
Mutant	−39.0	−46.4	0.0	–	5.7	176.6	23.6	1.8	0.0	0.0	641 (0.74)	158 (0.18)	72 (0.08)
1	−52.1	−59.3	4.6	11	4.1	179.3	23.4	0.0	14.6	3.7	738 (0.69)	224 (0.21)	110 (0.10)
2	−51.3	−43.6	23.1	2	24.1	162.6	22.3	3.6	140.5	15.4	693 (0.51)	165 (0.12)	499 (0.37)
3	−49.0	−56.9	3.3	3	5.2	178.9	23.5	0.1	11.2	2.6	748 (0.72)	195 (0.19)	101 (0.10)
4	−46.4	−42.2	18.0	3	13.9	174.0	23.0	4.9	85.8	11.8	494 (0.49)	222 (0.22)	296 (0.29)
5	−46.2	−38.9	21.6	4	22.8	133.1	23.5	10.7	153.1	12.7	394 (0.46)	93 (0.11)	362 (0.43)
6	−45.4	−40.0	19.6	1	15.7	174.4	23.0	5.3	95.1	13.0	497 (0.48)	213 (0.20)	331 (0.32)
7	−45.1	−39.8	22.7	1	23.3	162.5	23.5	3.2	141.1	15.0	620 (0.54)	114 (0.10)	412 (0.36)
8	−44.2	−39.7	18.7	1	19.6	160.6	22.9	5.7	98.3	10.6	526 (0.51)	183 (0.18)	328 (0.32)
9	−43.4	−42.5	16.8	3	23.7	153.0	23.8	21.7	150.5	1.8	489 (0.54)	195 (0.22)	212 (0.24)
10	−43.4	−37.4	20.8	3	23.1	126.6	23.4	10.7	157.7	11.0	407 (0.48)	100 (0.12)	331 (0.40)

[a] Statistics listed are for the lowest energy docking in each cluster. Total number of dockings that passed the energy cutoff (−30.0 kcal/mol) in this experiment was 184.
[b] See Table 5.
[c] A simple correction for solvation effects was made (see Materials and methods).
[d] See Table 5.
[e] See Table 5.
[f] PROBE:75:C to TARGET:48:NZ distance.
[g] Rotation necessary to superimpose the docking onto the target and the distance between centers of the two molecules.
[h] See Table 3.
[i] Rotation necessary to superimpose the docking onto the reference probe and the distance between centers of the two molecules.
[j] See Table 3.

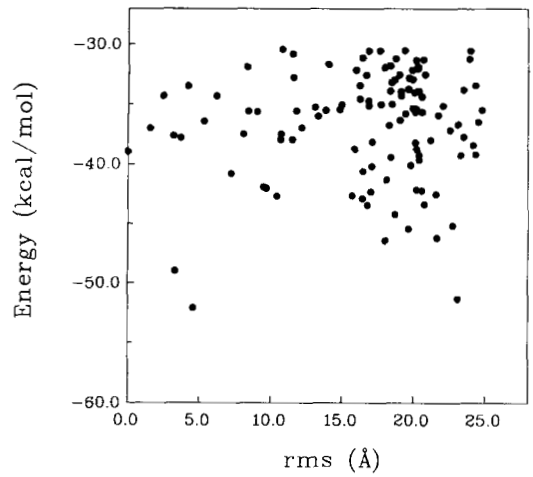

Fig. 5. Reconstruction of diubiquitin from "mutant" ubiquitin, judged by RMS deviation from a reference structure. Forty thousand separate docking starts with two copies of our modified ubiquitin molecule gave rise to 184 dockings with interaction energies below −30.0 kcal/mol. These were separated into 110 clusters with the three lowest energy clusters containing a total of 16 dockings. The lowest energy member of each of the 110 clusters is shown in this figure. The point with an RMS value of 0.0 represents the minimized reference configuration.

are within 2 Å RMS of our reference structure. Furthermore, the top 10 clusters are all represented by dockings of lower energy than the reference structure. However, by exploiting additional biochemical and structural information, as discussed below, the ambiguity can be removed. We describe several analyses which together clearly indicate that clusters 1 and 3 are essentially correct dockings. It is thus possible to predict clearly a diubiquitin-like dimer from two copies of the modified monomer, using information derived independently of the diubiquitin structure.

We applied Eisenberg and McLachlan's (1986) correction for solvation effects (not considered in our current energy calculation) to the calculated interaction energies and this improved the relative ranking of the top 10 clusters such that only clusters 1 and 3 were of lower energy than the reference structure (Table 6). This is a reflection of the nature and extent of the buried surface in each of the dockings (Table 6). Janin et al. (1988) and Miller (1989) have compiled a detailed summary of the nature of the accessible surfaces of both monomers and polymers. Both ubiquitin and diubiquitin fit the description given by these authors regarding the nature of accessible surface area as well as that of the dimer interface. Of the 18 dimer interfaces studied by these authors (Janin et al., 1988), none were more than approximately 22% charged or 30% polar. The information regarding charged interface surface area is compelling. Although both diubiquitin and our reference dimer fall within the boundaries outlined by these authors, all of the top 10 clusters, with the exception of clusters 1 and 3, have relatively high proportions of buried charged surfaces in the dimer interface (Table 6). The application of this type of information as a data filter is a considerable aid in the analysis of (the huge amount of) data obtained in these docking simulations. The potential function used in the current docking simulations does not consider solvent effects at all and these results clearly indicate the need for, and potential utility of, such a term in our energy calculation. On the other hand, the dramatic improvement we obtained by applying this solvation correction to our docking results is in contrast to the reports of other groups. Shoichet and Kuntz (1991) reported docking results of both bound and unbound molecules for three different systems. Applying a similar correction to the one we used (compare the ASP values of Eisenberg & McLachlan [1986] with those of Eisenberg et al. [1989]), they achieved no significant improvement in the relative energy rankings of any of their reported dockings. It is possible that these authors might have achieved greater success with this approach if they had used the parameter set employed in the present work. Alternatively, these discrepancies may indicate a lack of general applicability of this method of correcting for solvation effects. Detailed studies aimed at addressing this question are currently underway in this laboratory. Janin and co-workers (Cherfils et al., 1991, 1994) found no correlation between docking "correctness" and total buried surface area. These latter authors did not attempt to distinguish between various surface types in their calculations.

A mechanistic or functional analysis of the dockings also proved to be quite useful. Several other groups have derived distance constraints from a variety of nonenergetic information, and the application of such distance filters has been shown to simplify the analysis of a variety of docking results (Cherfils et al., 1991, 1994; Shoichet & Kuntz, 1991; Stoddard & Koshland, 1992). Because we were interested in a diubiquitin structure linked by an isopeptide bond between the C-terminus of the probe and Lys 48 of the target, we measured a representative distance for the top 10 clusters as well for two of our reference structures (Table 6). Because in all of our experiments we deleted residue 76, we measured the distance between PROBE:75:C and TARGET:48:NZ, in the reference structure(s) as well as the dockings, to determine the feasibility of isopeptide bond formation between the probe and the target (Table 6). Neglecting the possibility of large conformational shifts, isopeptide bond formation between the two molecules is only possible for clusters 1 and 3. Because the C-terminus of the probe as well as Lys 48 of the target are both flexible, we further explored the possibility of bond formation by running a systematic conformational search on these flexible regions of the molecules. A docking from an earlier experiment that was similar (0.7 Å RMS) to the low energy docking of cluster 2 had its C-terminus trimmed from Leu-Arg-Gly-Gly to Ala-Ala-Gly-Gly to simulate side-chain flexibility. We then systematically searched the accessible conformational space of these four ϕ–ψ pairs, as well as the four side-chain torsion angles of Lys 48 of the target, in 30° steps. With scaled down van der Waals radii, the closest conformer had a PROBE:75:C to TARGET:48:NZ distance of 9.0 Å. Using the native sequence (side chains of Leu 73 and Arg 74 fixed), we obtained no conformers in which this distance was less than 12 Å. In contrast, when we searched the four side-chain angles of Lys 48 and only the last ϕ angle in the C-terminus of the probe in our reference configuration, we obtained 408 conformers in which the distance of interest was between 2 and 4 Å.

We also examined these docking results for configurations representative of dimers that could be covalently linked via one of the other Lys residues of the target. Very few of the low energy dockings had pseudo-bond distances that would allow for covalent bond formation between the C-terminus of the probe and any of these other Lys residues of the target without ma-

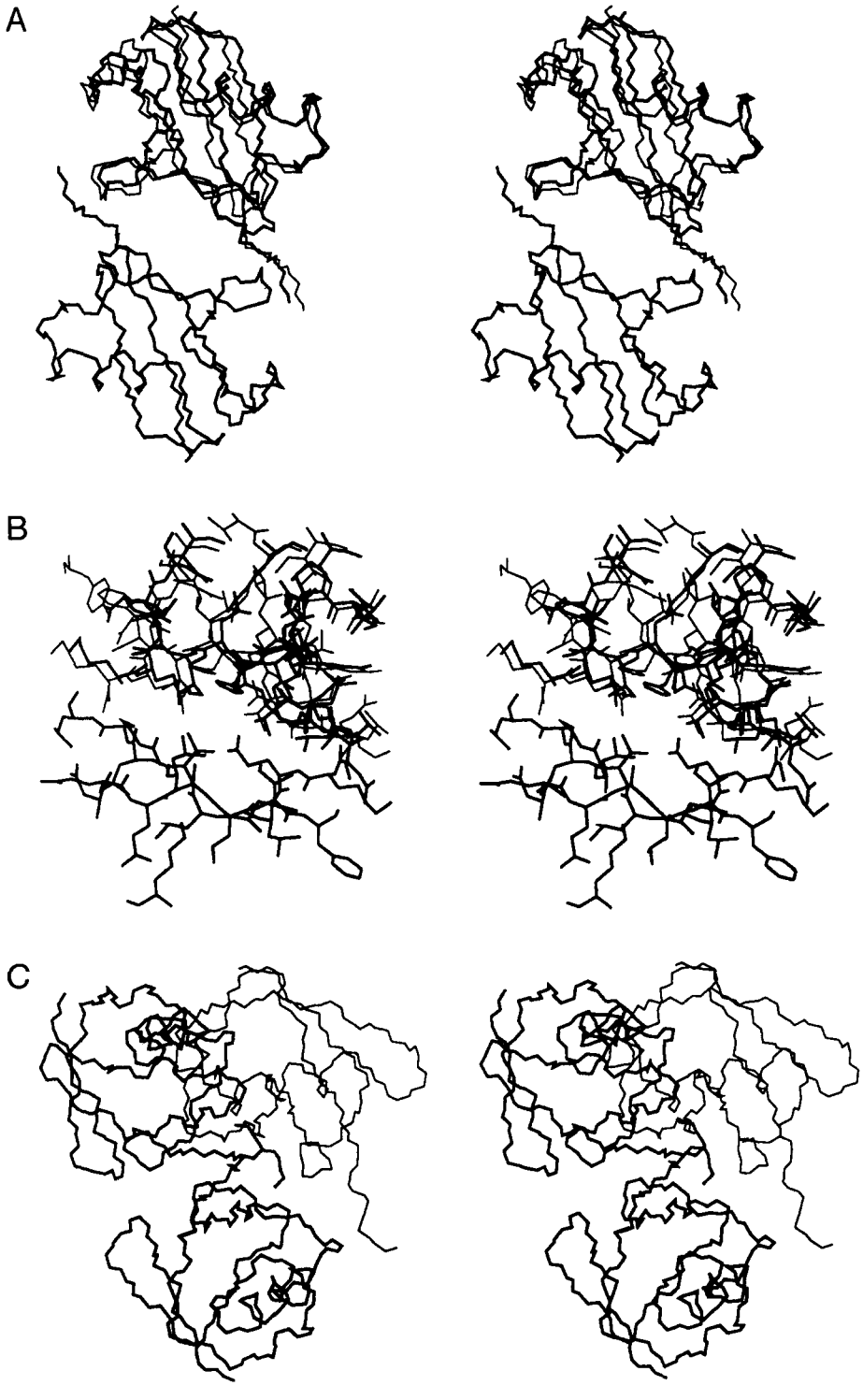

Fig. 6. Docking with monoubiquitin. Each stereo pair shows a target (lower; thick line) as well as a superposed reference (upper; thick lines) and docking (upper; thin lines). **A:** N, C, CA atoms of the lowest energy docking in cluster 1 (interaction energy = −52.1 kcal/mol; see Table 6). **B:** Details of the interface of the docking shown in A. In this figure, all heavy atoms of all residues containing an atom within 12 Å of CB of the modified (Arg 42 → Ala 42) residue (in the reference probe) are shown. **C:** N, C, CA atoms of the lowest energy docking in cluster 2 (interaction energy = −51.3 kcal/mol; see Table 6).

jor conformational adjustments. The lowest such distance for cluster 2 (Table 6) was 13.1 Å for Lys 27 of the target. Cluster 8 (Table 6) had a pseudo-bond distance of 7.5 Å with Lys 6 of the target, and cluster 19 had an equivalent distance of 9.1 Å. With these two exceptions, the most energetically favorable dockings that meet this covalent constraint are those represented by clusters 1 and 3. The most likely covalent linkage for these dockings is quite clearly the same as that observed in the diubiquitin structure. With information of this sort available, this type of data filtering would be of obvious value in a *real* prediction situation, where a reference structure is not available.

That a homodimer will generally exhibit twofold symmetry, or at least pseudosymmetry, was first predicted by Monod et al. (1965) and is supported by the empirical work of Miller (1989)

and others (see references in Miller, 1989). We measured the rotation and translation necessary to superimpose our dockings onto the target molecule. Diubiquitin exhibits twofold pseudosymmetry and our reference dimer (modified) is similar (Table 6). Clusters 1 and 3 have much closer twofold pseudosymmetry than most of the other clusters (Table 6). We used a method similar to that described by Shoichet and Kuntz (1991) to measure the difference between the dockings and the reference probe in terms of rotation angle and translation distance. Table 6 shows that this measurement approximately parallels the ranking according to the RMS differences between the dockings and the reference probe.

One final point worthy of mention is our criterion of which structures are the same. All of our analyses have been based on the arbitrary assumption that structures within 2 Å RMS of each other are the same, whereas more distant structures are different. For complexes involving two large molecules, particularly if the complex is of relatively low affinity, this may be an unrealistically limiting criterion. Indeed, our preliminary findings in this area with both gradient and Monte Carlo minimization suggest that more distant (than 2 Å RMS) configurations often converge to the same minimum (results not shown). Also, the relationship between the diubiquitin and tetraubiquitin (see below) structures supports the idea that diubiquitin has considerable configurational adaptability in solution.

Docking – Tetraubiquitin

While the present manuscript was in preparation, the structure of tetraubiquitin was reported (Cook et al., 1994). In contrast to the previously reported diubiquitin structure that is the focus of the current work, the ubiquitin-ubiquitin interactions in tetraubiquitin allow for indefinite extension of the ubiquitin polymer along a twofold screw axis. As a docking problem the prediction of the ubiquitin-ubiquitin configuration observed in the tetraubiquitin structure is much more difficult because a given ubiquitin monomer interacts with more than one other ubiquitin monomer.

We were, of course, interested in reexamining our results in light of this new information. A comparison of diubiquitin (only Gly 76 deleted; probe 1 in Table 3) and the appropriate dimer from the tetraubiquitin structure was most telling. With our potential function, we calculated interaction energies of −75.7 kcal/mol for diubiquitin and −9.7 kcal/mol for the tetraubiquitin dimer. The corresponding values corrected for solvation effects (see above) were −81.4 and −4.9 kcal/mol, respectively. Upon minimization, the diubiquitin structure shifted 0.7 Å RMS and the energy decreased slightly to −79.2 kcal/mol. Minimization of the tetraubiquitin dimer produced a more dramatic shift of 3.4 Å RMS and a new interaction energy of −33.4 kcal/mol. Solvation correction of these latter two interaction energies gave values of −85.0 and −28.1 kcal/mol, respectively. The interface area of the tetraubiquitin dimer is relatively small (595 $Å^2$) and the proportion of charged area is very high (45%) when compared to other dimers (Janin et al., 1988; Miller, 1989) as well as our reference diubiquitin configurations and dockings (Tables 3, 6).

Superimposing the target of the appropriate dimer from the tetraubiquitin structure onto the target in our docking simulation revealed that the probe from the new configuration extended, unfortunately, approximately 4 Å beyond the search space of our simulations. We would not, therefore, have found this configuration in our earlier docking experiments. Using a slightly larger search space (56-Å cube) placed so as to easily accommodate the new configuration (shifted 8 Å along one axis), as well as our earlier results, we constructed a reference "tetraubiquitin dimer" by superimposing our modified target and probe (see preceding section; Arg 42 → Ala and Gly 76 deleted) onto the appropriate halves of a Gly 76–Lys 48 isopeptide-linked pair from the tetraubiquitin structure. Rigid-body Monte Carlo minimization of this configuration resulted in a relatively large shift of 5.6 Å RMS. The original configuration had an interaction energy of +143.0 kcal/mol; minimization reduced this to −35.0 kcal/mol, so this reference structure did pass the (arbitrarily chosen) energy cutoff employed in our earlier simulations. The difference between the effects of minimization on this dimer and the native tetraubiquitin dimer (this dimer shifted 2.2 Å RMS more than the native dimer; see preceding paragraph) could not be ascribed to one or a few particular side-chain conformations or steric clashes. It is likely a reflection of the unsuitability of this interface for a simple monomer-monomer interaction. When we ran a docking simulation with this dimer in the new search space (see above) with 40,000 starts, 162 dockings passed the energy cutoff of −30.0 kcal/mol (results not shown). One relatively high energy docking was 4.2 Å RMS away from the minimized reference probe; no other dockings were within 11 Å RMS of this reference structure. On the other hand, several diubiquitin-like configurations *were* obtained, and some of these were among the lowest energy configurations observed. The most energetically favorable of these dockings was 3.5 Å RMS away from the minimized diubiquitin reference and had an interaction energy of −49.0 kcal/mol. The PROBE:75:C to TARGET:48:NZ distance of this docking was 5.2 Å, similar to that of the reference configuration (see above and Table 6). One of the eight dockings of lower energy (−50.2 kcal/mol) that were obtained in this simulation also had a favorable PROBE:75:C to TARGET:48:NZ distance (4.8 Å). This docking was 5.0 Å RMS away from the minimized diubiquitin reference configuration. The average PROBE:75:C to TARGET:48:NZ distance of the other seven lower energy dockings was 18.8 Å; the smallest was 14.8 Å. Without major conformational changes, covalent bond formation between the two molecules seems possible for only two of these low energy dockings.

This docking result suggested that, in the absence of further interactions, such as those observed in tetraubiquitin, and also, presumably, in higher order ubiquitin polymers, the dimer configuration observed in tetraubiquitin is not particularly favorable, at least according to our potential function. To explore this issue further, we ran a docking simulation in which we attempted to reassemble the two unmodified halves of a dimer taken from the tetraubiquitin structure (this experiment was analogous to our first two diubiquitin experiments with probe 1). In this case, very few dockings passed the −30.0-kcal/mol energy cutoff and none were within 15 Å RMS of the native or minimized reference probe. This result offers further support for the contention that, for a Gly 76–Lys 48 isopeptide-linked ubiquitin *dimer*, the monomer-monomer interaction observed in the tetraubiquitin structure is not particularly stable.

Taken together, our results indicate that a Gly 76–Lys 48 isopeptide-linked ubiquitin dimer can find more favorable interactions than those present between two adjacent monomers in the tetraubiquitin structure. The crystallographically ob-

served diubiquitin structure is an example of a more favorable ubiquitin–ubiquitin interaction and may represent the most favorable interaction for such a covalently linked pair. In discussing the tetraubiquitin structure, the authors (Cook et al., 1994) conclude that the diubiquitin structure probably represents the predominant form of ubiquitin in solution and that the tetraubiquitin configuration is adopted when a third monomer is ligated to diubiquitin. Our docking simulations, as well as our energy and surface area calculations, are consistent with this conclusion.

Docking—Summary

Reconstruction of a crystallographic complex is the standard test of a docking protocol. Several different methodologies, including the one employed here, have passed this test in studies with a variety of biochemical systems (references cited above). A more rigorous and realistic test of a docking protocol is to attempt to reproduce the experimentally determined structure of a complex using the structures of the uncomplexed (i.e., native) components of the complex. This has also been achieved in a number of cases (references cited above).

The present work differs significantly from previous examples, however, in a number of ways. First, we are studying a system that, to our knowledge, has not been investigated by such methods. Although we are predicting answers that have already been determined experimentally, our primary interest is in the prediction of the currently unknown structures of complexes involving ubiquitin and enzymes of the ubiquitin conjugation pathway. The quality of the results reported here indicates to us that we may be able to make such predictions reliably. Second, ubiquitin interacts with a variety of nonhomologous enzymes (Finley & Chau, 1991; Hershko, 1991; Rechsteiner, 1991; Hershko & Ciechanover, 1992; Hochstrasser, 1992; Jentsch, 1992). Most docking studies have focussed on target–probe interactions that are very specific and of relatively high affinity (see, for example, Goodsell & Olson, 1990; Shoichet & Kuntz, 1991; Bacon & Moult, 1992; Hart & Read, 1992). We are encouraged by our ability to predict the structure of a complex involving this "indiscriminate" protein. Third, we have applied a variety of nonenergetic biochemical information to the analysis of our docking results in a systematic, quantitative, and productive manner. Although most of the methods we applied have been reported previously, the variety of information we found to be applicable to this problem, as well as the extent to which these filters clarified the analysis of the results of our final docking simulation, is particularly encouraging. Fourth, most docking studies have investigated noncovalent complexes. It is unknown what part, if any, noncovalent intermolecular interactions play in the formation or stabilization of covalent ubiquitin complexes. As Shoichet and Kuntz (1991) have observed previously, the existence of a covalent bond between the two components of a complex can potentially complicate, as well as simplify, a docking study. Our docking simulations were performed without consideration of covalent bond formation between the two ubiquitin subunits (except, in one case, for filtering the docking results). The results reported here suggest that noncovalent intermolecular interactions are important for the formation and/or stabilization of the crystallographically observed diubiquitin complex.

It might be argued that ubiquitin self-associates too weakly to provide a good system for docking studies. Nonetheless, it is not surprising that the noncovalent affinity is low in a complex with a covalent bond, because evolution will only proceed to the point that there is a moderate energy stabilizing the desired configuration(s). On the other hand, a requirement for specificity means that the energy *difference* between the desired configuration(s) and all other possibilities must be large compared to kT. The success of a docking experiment depends on the discrimination of energy differences of this size and not on absolute binding energies.

Another aspect of the present work is the development of general docking strategies. In this respect, the experiments reported here serve several purposes. First, although side-chain flexibility is crucial, at least in some cases, to successful docking, we report several more examples of the effectiveness of a relatively crude approximation of this flexibility, residue truncation. Second, the consideration of only noncovalent interactions during docking can lead to correct predictions with a covalent complex. Third, and perhaps most important, the consideration of diverse chemical and biochemical information can dramatically clarify the results obtained from docking simulations. Fourth, we have seen that certain modifications of our docking procedure, such as the incorporation of a surface burial term and a different minimization scheme, could increase the power of that procedure.

We have discussed the limitations imposed by rigid-body docking and the difficulty of allowing for major conformational changes during docking simulations. Side-chain flexibility, on the other hand, can be modeled during docking simulations. In the current study, we chose to deal with the flexibility of certain critical residues in two ways—by systematically searching the accessible conformational states of a docked complex and by truncating relatively disordered residues during docking. Conformational searching proved to be a powerful way of incorporating biochemical information into the analysis of our docking results. Truncation, obviously, is a radical approximation of flexibility and by no means ideal, especially when the residue of interest is part of the intermolecular interface involved in the docking study. When the intermolecular interaction is between two proteins and involves a large interface the truncation of one or two side chains may remove a prohibitive steric clash without otherwise affecting the association. A better approximation would be to include a limited rotamer library of flexible side chains that could be sampled during the docking simulation. The application of the dead-end elimination theorem to the prediction of side-chain conformation has recently been described (Desmet et al., 1992), and its incorporation into a docking protocol has been reported (Leach, 1994). This method is also based on a library of allowed side-chain rotamers. The judicious implementation of some type of discrete-sampling approach, based on a user-defined rotamer library, to address the problem of side-chain flexibility in docking seems to be computationally feasible at this time, and we plan to incorporate such an improvement in our method.

The ideal docking experiment would search all possible configurations of the complex of interest and pick the correct one to be the one of lowest energy. Furthermore, this conclusion would be arrived at without considering any additional information (e.g., binding or mutation studies, chemical modifications). Current methods do not allow for this ideal experiment due to a variety of limitations.

Drug design is one of the common applications of docking simulations. Consideration of a typical drug-design scenario, however, leads one to the conclusion that such a powerful method is not strictly necessary (although it is, of course, desirable). Simply speaking, in this scenario, the investigators will have a target structure derived from either experiment or calculation, several structurally related ligands that exhibit a wide range of affinities for the target, and some information regarding the nature of the site of interaction (from, for example, competition or mutation studies). This type of information can be incorporated into a docking study to greatly reduce the configurational space that must be searched. This in turn will allow for a more exhaustive search of the smaller space and will increase the chances of determining the correct binding mode(s). Alternatively, such information can be applied as a filter to reject some of the data obtained in an unrestricted docking simulation. Variations of these approaches have been reported by several other groups (references cited above). We have successfully applied both of these approaches in the present studies of the diubiquitin system.

Conclusion

Upon consideration of the limitations and approximations involved in current docking simulations, indeed in simulations of biomolecules in general, surprise at the quality of the results obtained in many of these simulations is, perhaps, justified. Although the state of the art of simulations continues to evolve toward a truer representation of reality, the disparity that exists between current ideals and implementations will undoubtedly persist for several years. Irrespective of this, many workers continue to achieve success in the field.

In the current work we have applied our particular implementation of a method to solve the docking problem to a new and challenging biochemical system. Consideration of biochemical and structural information derived from a variety of sources and the application of such to both experimental design and the analysis of our results has allowed us to generate and correctly score diubiquitin-like dimer configurations using the two halves of diubiquitin, two copies of one of the halves of diubiquitin, as well as two copies of a modified form of the uncomplexed ubiquitin monomer. Docking results, as well as the results of surface area and energy calculations, are consistent with the observation of distinct configurations for a ubiquitin dimer and tetramer. The monomer-monomer interaction observed in tetraubiquitin is relatively unfavorable for a simple covalently linked ubiquitin dimer. Our ability to predict the crystallographically observed dimer configuration supports the idea that this structure represents the biologically relevant dimer configuration. Future work on this system will focus on the prediction of complexes of ubiquitin with enzymes of the ubiquitin conjugation pathway.

Finally, we briefly discuss the evolution of our docking method. One limitation of current concern is the inefficiency of Monte Carlo minimization in getting to the bottom of local minima. In the near future our method will be modified to include a two-stage minimization scheme involving an initial stage of Monte Carlo search followed by a final stage of gradient minimization. Our preliminary investigations in this area indicate that the clusters will become much tighter, reducing the complexity of the results obtained from docking simulations. Also, although we consider our docking simulations to have been successful in the three systems described here, as well as those described elsewhere (Hart & Read, 1992, 1994), the data presented clearly show that incorporation of a surface burial term in our potential function would help to clarify the docking results obtained, at least in this case. This observation is not surprising and we plan to incorporate a term to account for solvation effects in the next version of BOXSEARCH. Finally, we will be introducing methods for more accurately modeling flexibility as well as for incorporating relevant biochemical and structural information into our simulations.

Acknowledgments

This research was supported by the Medical Research Council of Canada and in part by an International Research Scholarship to R.J.R. from the Howard Hughes Medical Institute. R.J.R. is a Scholar of the Alberta Heritage Foundation for Medical Research. M.D.C. was the holder of a Natural Sciences and Engineering Research Council of Canada postgraduate scholarship while on leave from SynPhar Laboratories, Inc., Edmonton. We thank Mike Ellison for bringing this challenging system to our attention and for helpful and stimulating discussions throughout the course of this work. We also thank Leigh Willard for help with VADAR, Stephen Evans for help with GRAPH, Stephen Ness for help with the development of some of the analysis tools, and the staff of the Biosym Response Center, particularly Fora Chan and Marvin Waldman, for help with various aspects of INSIGHT.

References

Bacon DJ, Moult J. 1992. Docking by least-squares fitting of molecular surface patterns. *J Mol Biol 225*:849–858.

Bass MB, Hopkins DF, Jaquysh WAN, Ornstein RL. 1992. A method for determining the positions of polar hydrogens added to a protein structure that maximizes protein hydrogen bonding. *Proteins Struct Funct Genet 12*:266–277.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Chau V, Tobias JW, Bachmair A, Marriott D, Ecker DJ, Gonda DK, Varshavsky A. 1989. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science 243*:1576–1583.

Chen Z, Pickart CM. 1990. A 25-kilodalton ubiquitin carrier protein (E2) catalyzes multiubiquitin chain synthesis via lysine 48 of ubiquitin. *J Biol Chem 265*:21835–21842.

Cherfils J, Bizebard T, Knossow M, Janin J. 1994. Rigid-body docking with mutant constraints of influenza hemagglutinin with antibody HC19. *Proteins Struct Funct Genet 18*:8–18.

Cherfils J, Duquerroy S, Janin J. 1991. Protein–protein recognition analyzed by docking simulation. *Proteins Struct Funct Genet 11*:271–280.

Cherfils J, Janin J. 1993. Protein docking algorithms: Simulating molecular recognition. *Curr Opin Struct Biol 3*:265–269.

Cook WJ, Jeffrey LC, Carson M, Chen Z, Pickart CM. 1992a. Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). *J Biol Chem 267*:16467–16471.

Cook WJ, Jeffrey LC, Kasperek E, Pickart CM. 1994. Structure of tetraubiquitin shows how multiubiquitin chains can be formed. *J Mol Biol 236*:601–609.

Cook WJ, Jeffrey LC, Sullivan ML, Vierstra RD. 1992b. Three-dimensional structure of a ubiquitin-conjugating enzyme (E2). *J Biol Chem 267*: 15116–15121.

Cook WJ, Jeffrey LC, Xu Y, Chau V. 1993. Tertiary structures of class I ubiquitin-conjugating enzymes are highly conserved: Crystal structure of yeast UBC4. *Biochemistry 32*:13809–13817.

Desmet J, De Maeyer M, Hazes B, Lasters I. 1992. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature 356*: 539–542.

Deveraux Q, Ustrell V, Pickart C, Rechsteiner M. 1994. A 26 S protease subunit that binds ubiquitin conjugates. *J Biol Chem 269*:7059–7061.

Ecker DJ, Butt TR, Marsh J, Sternberg EJ, Margolis N, Monia BP, Jonnalagadda S, Khan MI, Weber PL, Mueller L, Crooke ST. 1987. Gene synthesis, expression, structures, and functional activities site-specific mutants of ubiquitin. *J Biol Chem 262*:14213–14221.

Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. *Nature 319*:199–203.
Eisenberg D, Wesson M, Yamashita M. 1989. Interpretation of protein folding and binding with atomic solvation parameters. *Chem Scr 29A*:217–221.
Evans SV. 1993. Setor: Hardware lighted three-dimensional solid model representations of macromolecules. *J Mol Graphics 11*:134–138.
Finley D, Chau V. 1991. Ubiquitination. *Annu Rev Cell Biol 7*:25–69.
Goodsell DS, Olson AJ. 1990. Automated docking of substrates to proteins by simulated annealing. *Proteins Struct Funct Genet 8*:195–202.
Gregori L, Marriott D, West CM, Chau V. 1985. Specific recognition of calmodulin from *Dictyostelium discoideum* by the ATP,ubiquitin-dependent degradative pathway. *J Biol Chem 260*:5232–5235.
Gregori L, Poosch MS, Cousins G, Chau V. 1990. A uniform isopeptide-linked multiubiquitin chain is sufficient to target substrate for degradation in ubiquitin-mediated proteolysis. *J Biol Chem 265*:8354–8357.
Hart TN, Read RJ. 1992. A multiple-start Monte Carlo docking method. *Proteins Struct Funct Genet 13*:206–222.
Hart TN, Read RJ. 1994. Multiple-start Monte Carlo docking of flexible ligands. In: Merz KM Jr, LeGrand SM, eds. *The protein folding problem and tertiary structure prediction*. Boston: Birkhäuser. pp 71–108.
Hershko A. 1991. The ubiquitin pathway for protein degradation. *Trends Biochem Sci 16*:265–268.
Hershko A, Ciechanover A. 1992. The ubiquitin system for protein degradation. *Annu Rev Biochem 61*:761–807.
Hershko A, Heller H. 1985. Occurrence of a polyubiquitin structure in ubiquitin-protein conjugates. *Biochem Biophys Res Commun 128*:1079–1086.
Hill CP, Johnston NL, Cohen RE. 1993. Crystal structure of a ubiquitin-dependent degradation substrate: A three-disulfide form of lysozyme. *Proc Natl Acad Sci USA 90*:4136–4140.
Hochstrasser M. 1992. Ubiquitin and intracellular protein degradation. *Curr Opin Cell Biol 4*:1024–1031.
Janin J, Miller S, Chothia C. 1988. Surface, subunit interfaces and interior of oligomeric proteins. *J Mol Biol 204*:155–164.
Jentsch S. 1992. The ubiquitin-conjugation system. *Annu Rev Genet 26*:179–207.
Kuntz ID, Meng EC, Shoichet BK. 1994. Structure-based molecular design. *Acc Chem Res 27*:117–123.
Leach AR. 1994. Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol 235*:345–356.
Miller S. 1989. The structure of interfaces between subunits of dimeric and tetrameric proteins. *Protein Eng 3*:77–83.
Monod J, Wyman J, Changeux JP. 1965. On the nature of allosteric transitions: A plausible model. *J Mol Biol 12*:88–117.
Ozkaynak E, Finley D, Varshavsky A. 1984. The yeast ubiquitin gene: Head-to-tail repeats encoding a polyubiquitin precursor protein. *Nature 312*:663–666.
Rao ST, Rossmann MG. 1973. Comparison of super-secondary structures in proteins. *J Mol Biol 76*:241–256.
Rechsteiner M. 1991. Natural substrates of the ubiquitin proteolytic pathway. *Cell 66*:615–618.
Richmond TJ. 1984. Solvent accessible surface area and excluded volume in proteins. *J Mol Biol 178*:63–89.
Shoichet BK, Kuntz ID. 1991. Protein docking and complementarity. *J Mol Biol 221*:327–346.
Silver ET, Gwozd TJ, Ptak C, Goebl M, Ellison MJ. 1992. A chimeric ubiquitin conjugating enzyme that combines the cell cycle properties of CDC34 (UBC3) and the DNA repair properties of RAD6 (UBC6): Implications for the structure, function and evolution of the E2s. *EMBO J 11*:3091–3098.
Sokolik CW, Cohen RE. 1992. Ubiquitin conjugation to cytochromes *c*. *J Biol Chem 267*:1067–1071.
Stoddard BL, Koshland DE Jr. 1992. Prediction of the structure of a receptor-protein complex using a binary docking method. *Nature 358*:774–776.
Varshavsky A. 1992. The N-end rule. *Cell 69*:725–735.
Vierstra RD, Langan SM, Schaller GE. 1986. Complete amino acid sequence of ubiquitin from the higher plant *Avena sativa*. *Biochemistry 25*:3105–3108.
Vijay-Kumar S, Bugg CE, Cook WJ. 1987. Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol 194*:531–544.
Wilkinson KD. 1988. Purification and structural properties of ubiquitin. In: Rechsteiner M, ed. *Ubiquitin*. New York: Plenum Press. pp 5–38.
Wilkinson KD, Cox MJ, O'Connor LB, Shapira R. 1986. Structure and activities of a variant ubiquitin sequence from bakers' yeast. *Biochemistry 25*:4999–5004.