Protein Science (1997), *6*:1901–1909. Cambridge University Press. Printed in the USA.

Surface topology of Minibody by selective chemical modifications and mass spectrometry

FRANCESCA ZAPPACOSTA,[1,2] PAOLO INGALLINELLA,[1] ANDREA SCALONI,[1,3] ANTONELLO PESSI,[4] ELISABETTA BIANCHI,[4] MAURIZIO SOLLAZZO,[4] ANNA TRAMONTANO,[4] GENNARO MARINO,[2,5] AND PIERO PUCCI[1,2]

[1]Centro Internazionale di Servizi di Spettrometria di Massa, CNR–Università di Napoli, Via S. Pansini 5, 80131 Napoli, Italy.
[2]Dipartimento di Chimica Organica e Biologica, Università di Napoli "Federico II," Napoli, Italy.
[3]IABBAM, CNR, Napoli, Italy.
[4]IRBM P. Angeletti, Pomezia, Italy.
[5]CEINGE, Biotecnologie Avanzate scrl, Napoli, Italy.

(RECEIVED March 5, 1997; ACCEPTED May 29, 1997)

Abstract

The surface topology of the Minibody, a small de novo-designed β-protein, has been probed by a strategy that combines selective chemical modification with a variety of reagents and mass spectrometric analysis of the modified fragments. Under appropriate conditions, the susceptibility of individual residues primarily depends on their surface accessibility so that their relative reactivities can be correlated with their position in the tertiary structure of the protein. Moreover, this approach provides information on interacting residues, since intramolecular interactions might greatly affect the reactivity of individual side chains by altering their pKa values.

The results of this study indicate that, while overall the Minibody model is correct, the β-sheet formed by the N- and C-terminal segments is most likely distorted. This is also in agreement with previous results that were obtained using a similar approach where mass spectrometry was used to identify Minibody fragments from limited proteolysis (Zappacosta F, Pessi A, Bianchi E, Venturini S, Sollazzo M, Tramontano A, Marino G, Pucci P. 1996. Probing the tertiary structure of proteins by limited proteolysis and mass spectrometry: The case of Minibody. *Protein Sci 5*:802–813). The chemical modification approach, in combination with limited proteolysis procedures, can provide useful, albeit partial, structural information to complement simulation techniques. This is especially valuable when, as in the Minibody case, an NMR and/or X-ray structure cannot be obtained due to insufficient solubility of the molecule.

Keywords: mass spectrometry; Minibody; protein surface topology; selective chemical modifications

The Minibody is a de novo-designed 61-residue metal-binding protein consisting of a β-sheet scaffold based on the heavy-chain variable domain structure of a mouse immunoglobulin and containing two segments corresponding to its exposed hypervariable H1 and H2 loops (Pessi et al., 1993; Bianchi et al., 1994; Tramontano et al., 1994). The design was aimed at creating a small scaffold composed of well-characterized secondary structure elements and tolerant of sequence variability to display on its surface conformationally-constrained peptides whose conformation can be predicted (Tramontano et al., 1994). A filamentous phage Minibody library has been successfully used to select an inhibitor of human Interleukin-6 (Martin et al., 1994, 1996).

Circular dichroism spectroscopy, size exclusion chromatography, and metal-binding studies confirmed the general fold of the protein (Pessi et al., 1993). However, the detailed investigation of the three-dimensional structure by NMR spectroscopy and X-ray diffraction analysis was initially thwarted by the low solubility of the protein. Addition of three lysine residues to the C-terminus yielded a mutant, named MB1-K3 (Bianchi et al., 1994), soluble enough for NMR, albeit only at very low pH. This study provided data on the general conformation of the protein backbone but no details about the amino acid side chains (Bazzo et al., personal communication).

Altogether, the overall conformation of Minibody deduced from these data is consistent with the original design, which embodies three β-strands from each of the two β-sheets of the heavy chain variable domain of the immunoglobulin connected by five loops. Three of these are short segments linking strands 2–3, 4–5, and 5–6, respectively, and two are longer loops connecting strands 1–2 and 3–4, respectively, and corresponding to the H1 and H2 hypervariable regions of immunoglobulins (Pessi et al., 1993).

To gain further insights about the details of the protein structure in the physiological conditions in which it is active (Pessi et al., 1993; Martin et al., 1994), we used a strategy that integrates lim-

Reprint requests to: Prof. Piero Pucci, Centro Internazionale di Servizi di Spettrometria di Massa, CNR-Università di Napoli, Via S. Pansini 5, 80131 Napoli, Italy.

ited proteolysis experiments with the identification of the fragments released from the protein by electrospray mass spectrometry (ESMS) (Zappacosta et al., 1996). The rationale behind the method is that amino acid residues likely to be accessible to proteases under strictly controlled conditions are located within exposed and flexible loops or within the linking regions of the polypeptide chain between domains. Therefore, the identification of the preferred cleavage sites are strongly indicative of the exposed regions in the protein structure.

Complementary information was now obtained by an independent approach consisting of the selective chemical modification of specific residues located at the protein surface. Under conditions of limited chemical modifications, in fact, relative reactivities of specific residues depend primarily on their surface accessibility. Moreover, electrostatic interactions, hydrogen bonding, or other microenvironmental effects may influence the chemical properties of the amino acid side chains. Systematic modification experiments, performed using various chemical reagents with different specificities, followed by the identification of the modified amino acids, provide information about the exposed regions of the protein and possibly the side chains involved in intramolecular interactions.

Chemical modification reactions on proteins have long been employed in structure–function studies aimed at identifying active-site residues, studying chemical reactivity of specific amino acids, or modifying catalytic properties and proteolytic specificity (Glazer, 1976; Burnens et al., 1987). More recently, acetylation of lysine residues or modification of arginines were used to probe the surface topology of model proteins (Suckau et al., 1992; Glocker et al., 1994) or to investigate interacting regions in protein complexes (Steiner et al., 1991; Ohguro et al., 1994). The power of these methods is greatly improved when mass spectrometric techniques are used to determine both the precise number and location of the derived amino acid residues.

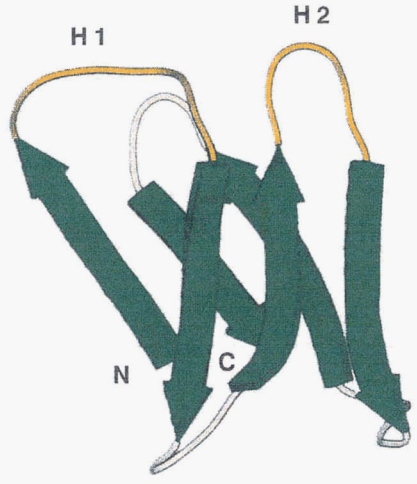
Here we show the results of a systematic study of the relative reactivities to different chemical reagents of different classes of amino acids of the Minibody such as lysines, arginines, and tyrosines. Our results correlate well with those previously provided by the limited proteolysis approach (Zappacosta et al., 1996), highlighting the potential of both procedures in providing useful information about the surface topography of a protein.

Results

Recombinant Minibody mutant MB1-K3 was expressed in *Escherichia coli*, purified and characterized as previously described (Pessi et al., 1993; Zappacosta et al., 1996). The amino acid sequence of the protein and its secondary structure are shown in Figure 1.

The surface topology of MB1-K3 was probed by selective chemical modifications of specific amino acids carried out by incubating aliquots of the protein with different modifying agents. This experimental approach needs to meet two important requirements. First, the native conformation of the protein should be stable under the conditions used for the modification reactions, avoiding any unfolding of the polypeptide chain, even partial or local. Second, the relative reactivity of individual amino acids should be solely affected by the three-dimensional structure of the protein or by the involvement of specific side chains in intramolecular interactions.

Special care was then taken to define the appropriate experimental conditions, in terms of pH, temperature, time of incubation, and stoichiometry of each modifying agent. Preliminary experiments were performed with each chemical reagent using different experimental conditions in order to maximize the stability of the protein tertiary structure and to limit the extent of the modification to the desired number of amino acid residues. In the analysis of the results, only mono- and di-modified species of MB1-K3 were taken into consideration, assuming that the introduction of further modifying groups might result in, or might have been due to, the alteration of the protein conformation. However, it should be emphasized that any modified MB1-K3 component may consist of a population of molecular species carrying the same number of modifying groups but located at different sites.




```
                  5                   10                  15
ALA-ASN-SER-GLN-ALA-THR-SER-GLY-PHE-THR-PHE-SER-HIS-PHE-TYR-
                              -------------H1------------
                  20                  25                  30
MET-GLU-TRP-VAL-ARG-GLY-GLY-GLU-TYR-ILE-ALA-ALA-SER-ARG-HIS-
                                                        ---
                  35                  40                  45
LYS-HIS-ASN-LYS-TYR-THR-THR-GLU-TYR-SER-ALA-SER-VAL-LYS-GLY-
--------H2---------
                  50                  55                  60
ARG-TYR-ILE-VAL-SER-ARG-ASP-THR-SER-GLN-SER-ILE-LEU-TYR-LEU-

GLN-LYS-LYS-LYS
```

Fig. 1. Amino acid sequence and secondary structure assignment of the Minibody (H1 and H2 designate the hypervariable loops).

Chemical modification of lysine residues

Acetylation of the lysine residues of MBI-K3 was accomplished by reaction with a two-fold molar excess of acetic anhydride for 10 min at pH 7.5. The incubation mixture was then fractionated by HPLC in order to separate the differently modified species.

Figure 2 shows the corresponding HPLC profile; individual peaks were collected and directly submitted to ESMS analysis. Measurement of the accurate molecular mass of each component indicated the extent of acetylation leading to the identification of the different protein species. Peak A exhibited a molecular mass of 7412.7 ± 0.8 Da, corresponding to unmodified MB1-K3 (theoretical value = 7412.3 Da), whereas fractions B and C yielded molecular masses of 7453.5 ± 0.5 Da and 7495.4 ± 0.7 Da respectively, with a mass increase of about 42 and 84 Da, respectively, and were then identified as the mono- and di-acetylated forms of MB1-K3.

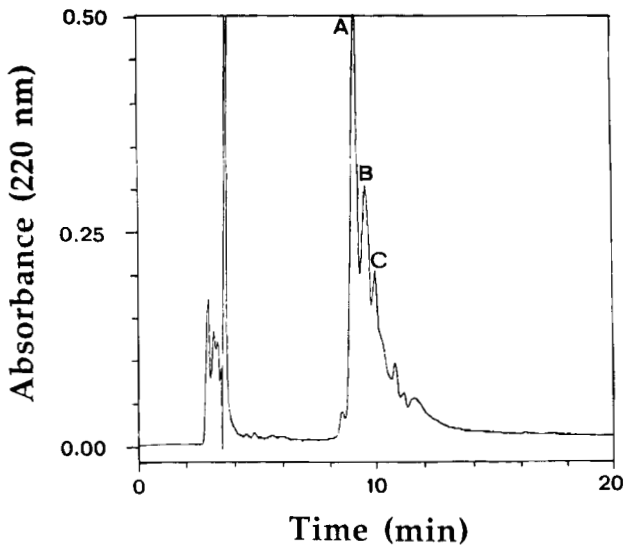


Fig. 2. Reverse phase HPLC profile of MB1-K3 modified by acetic anhydride. Peaks A, B and C correspond to the unmodified, mono- and di-modified protein species respectively as determined by ESMS analysis.

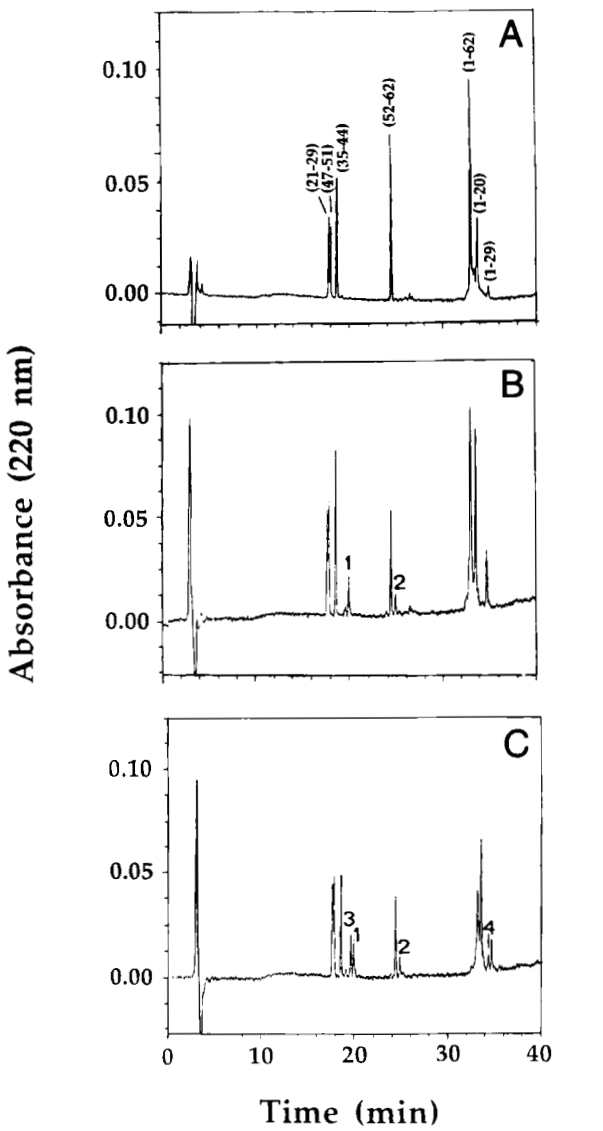


Fig. 3. HPLC separation of the tryptic digests from **(A)** unmodified, **(B)** mono-, and **(C)** di-acetylated forms of MB1-K3. The modified peptides identified by ESMS are indicated. Peaks corresponding to unmodified tryptic fragments are indicated in panel A.

In order to identify the lysine residues acetylated in each modified species, the mono- and the di-acetylated forms of MB1-K3 were individually digested with trypsin and the resulting peptide mixtures separated by HPLC. Figures 3B and C show the HPLC profiles obtained for the mono-acetylated and di-acetylated MB1-K3, respectively. As a comparison, the HPLC chromatogram of the tryptic digest of unmodified MB1-K3 is shown in Figure 3A; all the unmodified tryptic peptides were identified on the basis of their molecular mass. Peaks 1 and 2, present in both the mono- and di-acetylated digests, could not be assigned to any fragment within the MB1-K3 sequence. Mass spectrometric analysis of the corresponding fractions yielded molecular masses of 1833.0 ± 0.9 Da and 1465.2 ± 0.4 Da respectively, which were attributed to the peptides 30–44 and 52–63 both containing a single residue of N-ε-acetyl-lysine.

The modified lysine within the fragment 52–63 was immediately identified as Lys62; this residue was, in fact, not recognized by trypsin as cleavage site (Steiner et al., 1991; Zappacosta et al., 1994). Location of the N-ε-acetyl-lysine within the peptide 30–44 was not as straightforward and required amino acid sequence analysis. This fragment, in fact, is generated from an incomplete tryptic cleavage and contains two lysine residues at positions 31 and 34, both candidates for modification. Fraction 1 was then analyzed by automated Edman degradation that identified a single residue of N-ε-acetyl-lysine at position 34; no trace of modified Lys31 was observed.

Fractions 3 and 4, which occurred only in the di-acetylated form of MB1-K3 (Fig. 3C), were analogously analyzed by ESMS. On the basis of their molecular mass, these fragments were identified as mono-acetylated peptides 35–46 and 1–20, respectively. Both fragments contain a single modification site, Lys 44 and the protein N-terminal amino group.

These results clearly illustrate the differential reactivity displayed by the various amino groups present in MB1-K3. Lys34 and Lys62 were found to be acetylated in both the mono- and di-modified forms of the protein, thus suggesting that they represent the most reactive residues in the native conformation of the protein. Acetylation of the N-terminus and Lys44 was only detected in the di-modified form of MB1-K3, indicating that these residues reacted with acetic anhydride only when modification of the more reactive lysines had already occurred. Acetylation of Lys31 was never observed under these experimental conditions, suggesting that the side chain of this residue is either shielded by the tertiary structure of the protein or involved in intramolecular interactions that change the nucleophilic character of the ε-NH_2 group.

No information could be obtained on the modification of Lys63 and Lys64 because they are part of the three-lysine tail and consequently the putative fragments generated by trypsin would be very small; however, data concerning these residues are of limited value as they belong to the C-terminal tail, which is known to be exposed and very flexible (Bazzo et al., personal communication).

Table 1. *Electrospray mass spectrometric analysis of the MB1-K3 molecular species following incubation with N-acetyl-imidazole and HPLC separation*

Fraction	Molecular Mass	Component	Expected Value
A	7412.3 ± 1.5	MB1-K3	7412.3
B	7454.4 ± 0.6	Ac-MB1-K3	7454.3
C	7496.2 ± 0.9	Ac_2-MB1-K3	7496.4

Chemical modification of tyrosine residues

The tyrosine side chain consists of two moieties, the hydroxyl group and the aromatic ring, with very different chemical properties and normally located in different microenvironment within the tertiary structure of proteins. Modification of tyrosines with two different reagents, each specific for one of these moieties, can give more detailed information on the positioning of these residues on the protein and we used both N-acetyl-imidazole, a reagent specific for O-acetylation of the phenolic hydroxyl group and tetranitromethane, which selectively produces 3-nitrotyrosine (Lundblad R.L. and Noyes C.N. 1984).

O-acetylation was achieved by reaction of a five-fold molar excess of N-acetyl-imidazole for 30 min. The reaction was carried out at pH 6.5 to selectively modify the phenolic hydroxyl groups and avoid undesired side reactions with amino groups. The reaction mixture was then separated by HPLC and the individual fractions identified by ESMS analysis, as described above and reported in Table 1. Peak A corresponds to unmodified MB1-K3, whereas fractions B and C represent the mono- and di-O-acetylated forms of MB1-K3, respectively. Relative peak intensities suggest that under these conditions most of the protein remains unmodified or incorporates a single acetyl group, the doubly modified component representing only a minor species.

Identification of the O-acetyl-Tyr residues in the modified protein was achieved following the procedure described above. The mono- and di-acetylated species were individually digested with trypsin and the resulting peptide mixtures fractionated by HPLC; individual components were identified by ESMS analysis. Due to the chemical instability of phenolic esters at basic pH, the tryptic digestion was carried out at pH 6.0 for a limited time to minimise the hydrolysis of the acetyl groups.

Figure 4A shows the HPLC analysis of the peptide mixture generated by the enzymatic hydrolysis of mono-acetylated MB1-K3. As compared to the native digest, shown in Figure 3A, three anomalous peaks were detected and mass analyzed (Fig. 4B). Fractions 1 and 2 yielded the same molecular mass (1189.1 ± 0.2 and 1188.6 ± 1.0 Da) corresponding to fragment 35–44 in which a single tyrosine had been acetylated. Peptide 35–44 contains two Tyr residues at positions 35 and 39; the occurrence of two separated HPLC peaks showing an identical mass value was then interpreted as due to the same peptide carrying the modification at different sites and generating two fragments with different chromatographic properties. Fraction 3 showed a molecular mass of 1337.1±0.6 Da which was assigned to the mono-acetylated fragment 52–62. A single tyrosine residue occurs within this sequence at position 59, clearly indicating that Tyr59 had been converted into O-acetyl-Tyr.

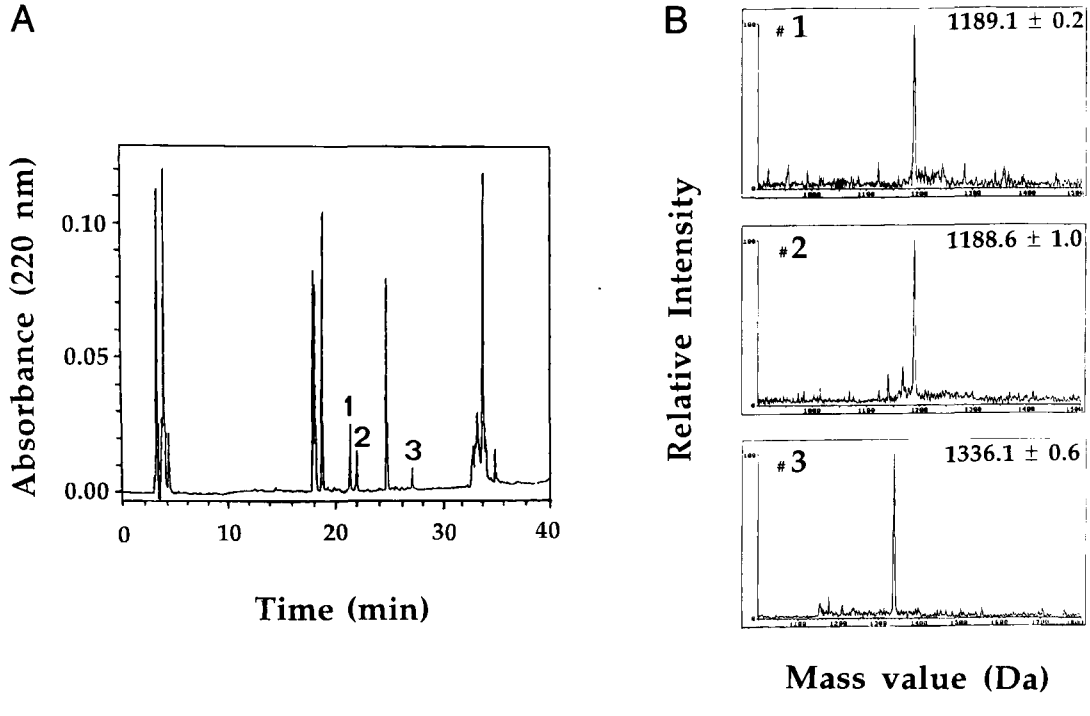


Fig. 4. A: HPLC separation of the tryptic digests from mono-O-acetylated MB1-K3. The modified peptides recognized by ESMS analysis are indicated; other peaks correspond to unmodified tryptic fragments. **B:** Transformed ESMS spectra of the modified fractions.

The same experimental approach was used to characterize the di-acetylated form of MB1-K3. Surprisingly, this analysis led to the identification of the same modification sites already observed in the mono-acetylated protein. Only Tyr35, Tyr39, and Tyr59, in fact, were found to be acetylated, indicating that even when MB1-K3 had been doubly modified, the acetyl groups were still distributed on the same three Tyr residues. Acetylation of the remaining three tyrosine residues, Tyr15, Tyr24, and Tyr47 was never observed.

Under mild conditions tetranitromethane is an efficient and specific reagent for the nitration of solvent-accessible tyrosines (Sokolovsky, 1966; Cuatrecasas, 1968); however, at higher concentrations side reactions such as oxidation or polymerization of the protein can easily occur (Lundblad & Noyes, 1984). The experimental conditions were then carefully chosen to control the reactivity of tetranitromethane and to constrain the reaction towards the selective nitration of tyrosine residues. Nitration of tyrosines was accomplished by 60 min reaction of a 10-fold molar excess of tetranitromethane at pH 7.5. Furthermore, since the reagent is dissolved in ethanol before being added to the MB1-K3 solution, the final concentration of the organic solvent was kept to a minimum thus ensuring the maintenance of the protein's native conformation.

Figure 5A shows the HPLC separation of the incubation mixture monitored both at 220 nm and 360 nm to take advantage of the specific absorbance of 3-nitrotyrosine at acidic pH. ES/MS analysis of the individual components confirmed the identity of peak A as native MB1-K3 and identified fractions B and C as the mono- and di-nitrated forms of the protein, respectively. The species eluted at higher retention times correspond to reaction by-products.

The elucidation of the nitrated tyrosine residues was carried out following the procedure outlined above. Figure 5B shows the HPLC profiles from the tryptic digestion of the mono-modified form of MB1-K3. The elution pattern was monitored at 220 and 360 nm allowing a rapid identification of four nitrated peptides, which were then analyzed by ESMS. On the basis of their mass value, fractions 1 and 4 were assigned to the peptides 47–51 and 52–62 with a mass increase of 45 Da indicating the incorporation of a single nitro group. As both fragments contain a single tyrosine residue, we concluded that Tyr47 and Tyr59 had been converted into 3-nitrotyrosine derivatives.

Mass spectrometric analysis of peaks 2 and 3 yielded the same molecular mass which was assigned to the mono-nitrated peptide 35–44. The situation here is identical to that observed during the O-acetylation procedure in that this peptide contains two tyrosines at positions 35 and 39 both candidates for the modification. Again, the difference in the retention times of these two fragments was attributed to a single modification event taking place at two different sites within the same peptide. However, the 3-nitrotyrosine derivative is stable under the Edman degradation conditions and the ambiguity could be solved by peptide sequencing of fractions 2 and 3 which showed that both peaks correspond to peptide 35–44, containing a Tyr–NO_2 either at position 35 (peak 2) or 39 (peak 3).

The same procedure was employed to identify the nitrated tyrosines in the di-modified form of MB1-K3. The HPLC profile of the corresponding tryptic digest, shown in Figure 5C, was essentially identical to that obtained for the mono-nitrated sample but for the occurrence of a further modified fragment. On the basis of its molecular mass, this peak (fraction 5) was assigned to fragment 21–29, in which Tyr24 had been transformed into 3-nitroTyr. All the other modified fractions yielded the same results obtained in the analysis of the mono-modified sample.

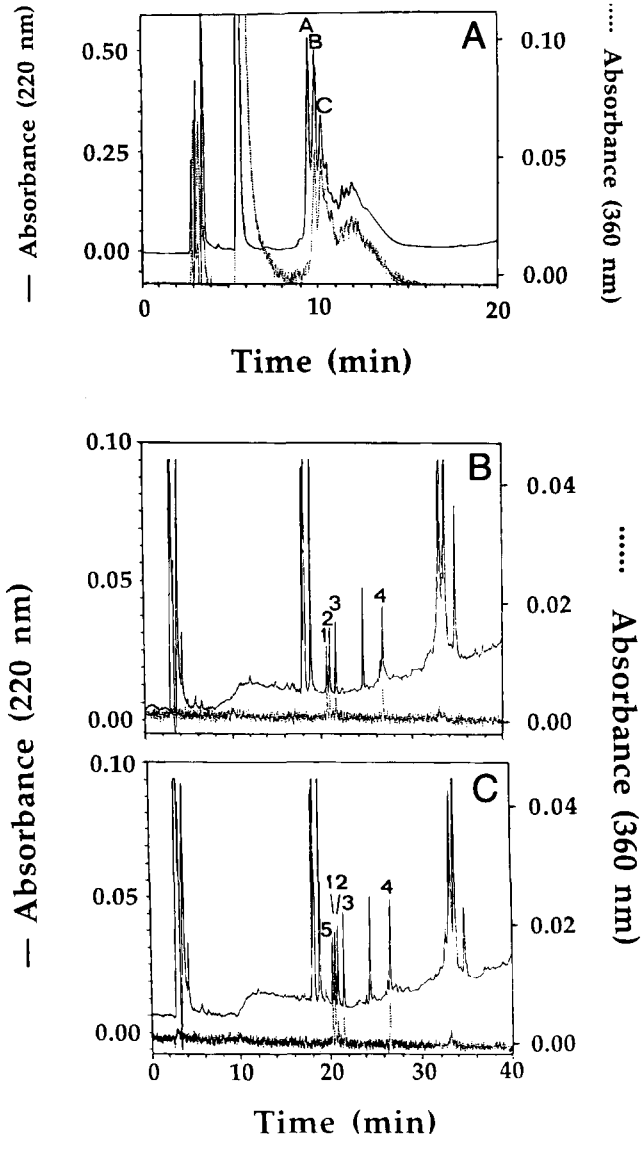


Fig. 5. A: HPLC profile of MB1-K3 modified with tetranitromethane. Peaks A, B, and C correspond to the unmodified, mono- and di-nitrated forms of the protein respectively. HPLC separation of the tryptic digests from the **(B)** mono and **(C)** di-nitrated MB1-K3 species. Modified peptides identified by ESMS are indicated.

These results clearly indicate that the six tyrosine residues present in the MB1-K3 molecule display a significantly different chemical reactivity. Tyr35, Tyr39, and Tyr59 are the most reactive residues, being derivatised by N-acetyl-imidazole and tetranitromethane in both the mono- and di-modified forms of MB1-K3. In contrast, Tyr24 and Tyr47 only reacted with tetranitromethane with the former having only been modified in the di-nitrated form of the protein, indicating a general lower chemical reactivity. It is worth noting that, under the experimental conditions used, Tyr15 was never found accessible to any of the two modifying agents.

Chemical modification of arginine residues

The chemical reactivity of arginine residues was investigated by incubation with a 70-fold molar excess of 1,2-cyclohexanedione

Table 2. *Electrospray mass spectrometric analysis of modified peptides from the tryptic digest of MB1-K3 following reaction with CHD*

Fraction	Molecular Mass	Peptide	Expected Value	Modification site
1	961.8 ± 0.5	45–51	961.9	Arg46
2	2351.9 ± 0.3	45–62	2351.4	Arg46 or Arg51
3	2026.6 ± 0.7	47–62	2026.1	Arg51
4	3382.9 ± 0.2	1–29	3383.5	Arg20

(CHD) for 30 min at pH 8.5. Modification of arginines with CHD mainly results in the formation of an N^7, N^8-(dihydroxy-1,2-cyclohexylidine)-derivative and a small percentage of the mono-hydroxy derivative with a corresponding mass increase of 112 Da and 94 Da, respectively (Suckau et al., 1992).

Under the experimental conditions used, the reaction yielded essentially the mono-modified form of MB1-K3 which co-eluted with the native protein. The HPLC separation of the reaction mixture showed a major component accompanied by a minor peak eluting at higher retention times. ESMS analysis demonstrated that the major component consisted of a quasi-equimolar amount of native and mono-modified MB1-K3, whereas the minor species was identified as the di-modified protein. The identification of the modified arginines in the most abundant component was then carried out on the mixture of the two forms.

Table 2 shows the mass spectrometric analysis of the corresponding tryptic digest; besides the expected tryptic peptides, four fragments were detected, each carrying a single modification group. Peptides 1–29, 45–51 and 47–62 contain a single arginine residue within their sequence, allowing the rapid identification of Arg20, Arg46, and Arg51 as the modification sites. As expected, the peptide bonds adjacent to the modified arginines were not recognized by trypsin. Fragment 45–62 originated from an incomplete tryptic cleavage and contains two arginine residues, indicating that either Arg46 or Arg51 had been converted into the corresponding N^7, N^8-(dihydroxy-1, 2-cyclohexylidine) derivative. Similar results were obtained when the di-modified form of MB1-K3 was analyzed; again three arginines out of four, namely Arg20, Arg46, and Arg51, were shown to react with CHD. In both cases Arg29, the fourth arginine residue of MB1-K3, was found unmodified.

According to Suckau et al. (1992), modification of protein guanidine groups with CHD can only occur if catalytic assistance by an intramolecular proton acceptor group is provided. On the basis of this consideration and unlike the other modification reactions, there is an inverse correlation between accessibility and chemical reactivity of arginines. The results described above lead to the conclusion that unmodified Arg29 is fully exposed to the solvent whereas the guanidinium groups of Arg20, Arg46, and Arg51 are involved in intramolecular electrostatic interaction with proton acceptor groups located in proximity of the N^7-guanidino function.

Discussion

We probed the surface topography of the Minibody by selectively modifying three classes of amino acids using various reagents with different specificity. Exposed residues usually display the highest chemical reactivity, although intramolecular interactions also affect the reactivity of some side chains by altering their pKa values (Suckau et al., 1992; Glocker et al., 1994).

Chemical modification conditions were carefully selected in order both to ensure maximum stability of the protein structure and to limit the extent of modification. If information about the tertiary structure of proteins has to be inferred by these experiments, it is crucial to maintain the native conformation under the experimental conditions used (Zappacosta et al., 1996). Any unfolding of the polypeptide chain, even partial or local, might result in the accumulation of misleading data. When a certain number of modifying groups is introduced within the protein structure, the possible perturbation of the subtle network of interactions that stabilizes the native conformation should be considered. Modification of amino acid side chains could result either in an unfavorable steric hindrance leading to local conformational changes or in the elimination of favorable interactions causing local distortions. Moreover, it is very difficult to predict whether or not the presence of exogenous moieties would disrupt the protein architecture.

Previous studies on model proteins reported that partial modification of lysines or arginines is compatible with the maintenance of native-like structures (Imoto et al., 1972; Glocker et al., 1994; Przybylski & Glocker, 1996). These findings are consistent with the observation that charged residues are usually exposed on the protein surface and mainly involved in protein–solvent interactions. Their modification is not expected to affect the overall conformation of the protein. However, when more than four arginines were modified in lysozyme, the ES analysis showed additional higher charge-state ions indicating gradual unfolding and denaturation of the protein (Loo et al., 1990; Suckau et al., 1992). Modification of arginines by CHD has been reported to occur solely at guanidino functions involved in intramolecular proton interactions (Suckau et al., 1992); elimination of these interactions could thus destabilize the protein structure, eventually causing unfolding. It is conceivable that greater problems may arise when the modification is addressed toward residues located, at least in part, in the interior of the protein molecule.

Two further considerations suggest the importance of limiting the extent of protein modification. First, the relative differences in chemical reactivity among the various amino acids will be enhanced at lower reagent excess. Second, once local unfolding has occurred, the accessibility of some side chains to further modifications might reflect the alteration of the native conformation and not the protein tertiary structure. This is even more important in the case of the Minibody which, due to low ΔG of unfolding (2.5 kcal/mol, Bianchi et al., 1994), might be easier to unfold. Consequently, only the mono- and di-modified forms of the protein were produced and considered for characterization.

Our results show that, within each class of amino acids, a certain number of residues are endowed with comparable reactivity, since mono- and di-modified components of MB1-K3 are found to consist of different molecular species carrying an equal number of modifying groups but located at different positions. A panel of chemical reagents with different specificity has been employed in this study with the aim of providing a systematic investigation of the protein surface accessibility. In particular, two modifying agents were used to probe the reactivity of tyrosine residues since the side chains of this amino acid consists of two distinctly susceptible moieties.

The results of modification experiments are summarized in Table 3. A relative ranking of chemical reactivity differences was derived on the basis of merely qualitative kinetic evaluation. Res-

Table 3. *Modification sites identified in MB1-K3 native structure following selective chemical derivatization under controlled conditions. The relative ranking of chemical reactivity differences was evaluated on merely qualitative kinetic basis.*

Modifying reagent	High reactivity residues	Low reactivity residues	Unmodified residues
Acetic Anhydride	Lys34, Lys62	NH_2-terminus, Lys44	Lys31
N-acetyl-imidazole	Tyr35, Tyr39, Tyr59		Tyr15, Tyr24, Tyr47
Tetranitromethane	Tyr35, Tyr39, Tyr47, Tyr59	Tyr24	Tyr15
CHD	Arg20, Arg46, Arg51		Arg29

idues labeled in both the mono- and di-modified forms of MB1-K3 were considered highly reactive, whereas the side chains labeled only in the di-modified species were endowed with a lower reactivity. These data correlate remarkably well with previous results obtained from limited proteolysis experiments (Zappacosta et al., 1996). As for the charged amino acids, Arg29, Lys34, and Lys62 displayed the highest reactivity and were then assumed to be the most accessible basic residues in MB1-K3. According to the predicted model, Arg29 and Lys34 are located within the H2 loop and their side chains are exposed to the solvent. The accessibility of Lys62 was also expected because this residue corresponds to the first lysine in the three-lysine tail added to the original Minibody.

Comparison of these results with the controlled proteolysis experiments shows that Arg29 and Lys62 are also preferred tryptic cleavage sites (Zappacosta et al., 1996). Interestingly, the results of the chemical modification experiments show that Lys34 has a high surface accessibility. This residue could not be cleaved by trypsin, and this was unexpected because the model predicted it to be exposed on the H2 loop, whose correct positioning had been inferred by several experiments (Pessi et al., 1993; Martin et al., 1994; Tramontano et al., 1994; Venturini et al., 1994). The lack of tryptic cleavage can now be attributed to conformational constraints of the surrounding polypeptide backbone, highlighting the complementarity of the chemical and enzymatic approaches.

The other basic residues are by far less accessible. In the design, Arg20, and Arg51 are both involved in intramolecular interactions and this is in good agreement with the CHD modification results. Arg51 is buried inside the protein core, whereas Arg20 is predicted to form a salt bridge with Glu23. Arg46, consistently with the absence of tryptic cleavage, was easily modified by CHD.

The N-terminal amino group and Lys44 showed a lower susceptibility, they are only modified after derivation of the more reactive groups. This suggests either a lower surface accessibility or a decreased nucleophilicity, or both and is consistent with the failure of trypsin in recognizing Lys44 (Zappacosta et al., 1996) and with the observation that the N-terminal segment of the protein is somewhat less accessible.

The results obtained for Lys31 with both limited proteolysis and chemical modification are again in agreement. This residue is neither modified by acetic anhydride nor cleaved by trypsin. This is, however, not consistent with the design and deserves some comments. In the model, Lys31 is located within the H2 loop, well exposed to the solvent. We tried to model its side chain so that it could be buried within the core and to identify a suitable partner for its ϵ-amino group, since unpaired charged residues are rarely found buried inside proteins. Our analysis of all possible conformations of this residues suggests two possibilities. First, it could be shielded from the solvent by the side chain of the solvent exposed Tyr35 forming hydrogen bonds to main-chain atoms of the surrounding region (whose precise identification is impossible from our data). The alternative possibility is to form a salt bridge with Asp 52, which would imply a distortion of the 46–52 β-strand; the distortion should, in any case, be limited to the carboxy-terminal region of the strand since our data for the surrounding residues are consistent with the model. It is difficult to discriminate between these two hypotheses using only the available data, and experiments are therefore in progress to substitute Lys31 with non-positively charged residues amenable to chemical modification and/or proteolysis. Limited proteolysis and selective chemical modifications of the mutants should help clarify the role played by electrostatic interactions in the positioning of the side chain of residue 31.

Investigation of the chemical reactivity of the six Tyr residues showed the high susceptibility of Tyr35, Tyr39, and Tyr59 towards both N-acetyl-imidazole and tetranitromethane, indicating that these residues are easily accessible. Again, a complete agreement with previous results from limited proteolysis experiments exists in that these three tyrosines were identified as primary proteolytic cleavage sites and assumed to be entirely exposed to the solvent. In particular, the accessibility of Tyr59 was attributed to the distortion of the protein structure caused by the too close proximity of the bulky side chain of Trp18 (Zappacosta et al., 1996).

In contrast, Tyr24 and Tyr47 were only modified by tetranitromethane with the former displaying a lower chemical reactivity. These results can be accounted for by a number of reasons; however, the most likely interpretation is that these residues are only partially accessible to the reagents and their phenolic hydroxyl groups are hydrogen bonded and then impaired in their reaction with N-acetyl imidazole. It should be noted that no data on these tyrosines could be obtained by protease digestions. Finally, Tyr15 was never modified under the conditions used, consistently with the presence of a compact and well-packed β-sheet formed by strands 2, 3, and 4.

The results of this study clearly demonstrate that selective chemical modification experiments can be designed to probe the surface accessibility of proteins and to gain information on their tertiary structure, provided that a systematic investigation using a variety of reagents is performed. In this respect, the employment of the largest possible number of modifying agents must be encouraged. For example, it has been recently reported that the oxidation of methionine residues by hydrogen peroxide provided some structural insights into the conformation of recombinant human stem cell factor (Hsu et al., 1996).

We would like to stress that the choice of the appropriate reagents must fulfill two main requirements. First, the modification reaction has to be carried out under conditions that ensure main-

tenance the protein native conformation. Second, the protein derivative must be stable enough to allow the identification of the modified residue(s). Several attempts to modify the histidine residues of Minibody with diethylpyrocarbonate were unsuccessful because the labeled histidine(s) was (were) converted back into the unmodified amino acid(s) under the analytical conditions used.

The chemical modification approach is complementary to the enzymatic procedure in providing information on the positioning of the amino acid side chains. Both techniques have the further advantage of providing data under physiological conditions, only requiring very limited amounts of material. This combined approach can be usefully extended to localize functional regions of the protein exposed to the solvent that become protected upon binding to other proteins or ligands, and possibly to monitor major conformational changes, for example, those occurring in the conversion of a native conformation into a stable folding intermediate.

Materials and methods

Materials

Trypsin, tetranitromethane, N-acetyl-imidazole, 1,2-cycloethanedione, and acetic anhydride were purchased from Sigma. All other reagents and solvents were HPLC-grade from Carlo Erba.

Expression of the Minibody gene and protein purification

The Minibody mutant MB1-K3 was obtained by recombinant DNA methodology (Bianchi et al. 1994) and purified as previously described (Bianchi et al., 1993). The purified protein was desalted by adsorption on a C-18 reverse phase matrix (Baker, Phillipsburg, NJ) and eluted with HFIP (Aldrich, Milwaukee, WI). After lyophilization, the polypeptide was stored at −20 °C and resuspended in the appropriate buffer before use.

Chemical modification experiments

Modification of specific residues of MB1-K3 was carried out by incubating aliquots of the protein (50 μg) with the appropriate reagent in 250 μL final volume. Acetylation of lysines and the N-terminal amino group was performed in 50 mM ammonium bicarbonate buffer, pH 7.5, at 25 °C for 10 min using a two-fold molar excess of acetic anhydride over the number of amino groups.

Modification of tyrosine residues was accomplished either by reaction with N-acetyl-imidazole or with tetranitromethane. Acetylation of phenolic hydroxyl groups was carried out in 50mM MES, pH 6.5, at 25 °C for 30 min, using a five-fold excess of reagent over the number of tyrosine residues. Nitration of the aromatic rings was performed in 25 mM sodium phosphate, pH 7.5, at 0 °C in the dark for 1 h with a 10-fold molar excess of tetranitromethane in ethanol over the number of tyrosine residues.

Modification of arginine residues was performed by reaction with 1,2-cyclohexanedione (CHD) in 50 mM sodium borate, pH 8.5, at 37 °C for 30 min using a 70-fold molar excess of reagent over the number of arginine residues.

All the reactions were stopped by acidification of the incubation mixtures with TFA 1% and the samples were directly injected onto a Vydac C18 column (250 × 4.6 mm, 5 μm, 300 Å pore size). The differently modified species of MB1-K3 were eluted by means of a linear gradient from 31% to 45% of acetonitrile in TFA 0.1% over 16 min; elution was monitored at 220 and/or at 360 nm. Individual fractions were collected and analyzed by ESMS.

Enzymatic hydrolysis of modified MB1-K3 and peptide separation

The differently modified samples of MB1-K3 were digested with trypsin in either 0.4% ammonium bicarbonate, pH 8.5, or 0.4% ammonium acetate, pH 6.0, at 37 °C for 2 h, using an enzyme: substrate ratio of 1:50 (w/w). Peptide mixtures were fractionated by RP-HPLC using a Vydac C18 column (250 × 4.6 mm, 5 μm, 300 Å pore size); peptides were eluted by means of a linear gradient from 5 to 50% of acetonitrile in TFA 0.1% over 35 min; elution was monitored at 220 and, in some cases, at 360 nm. Individual fractions were collected and analyzed by ESMS.

Peptide sequencing

Peptide sequencing of chemically modified fragments was performed by using a Perkin Elmer Applied Biosystem 477A pulsed-liquid protein sequencer equipped with a Perkin Elmer Applied Biosystem 120A HPLC apparatus for the PTH-amino acid identification. PTH-3-nitro-tyrosine and PTH-N-ε-acetyl-lysine standards were previously analyzed and their relative retention times compared with those of the unknown peaks present at the cycles where PTH–Tyr or PTH–Lys were expected.

Mass spectrometry

Differently modified species of MB1-K3 and their proteolytic fragments were analyzed by electrospray mass spectrometry using either a Bio-Q triple quadrupole mass spectrometer or a PLATFORM (Micromass) single quadrupole instrument both equipped with an electrospray ion source. Samples were directly injected into the ion source (kept at 60 °C) via a loop injection at a flow rate of 10 μL/min. Data were acquired and elaborated using the MASS-LYNX program (Micromass). Mass calibration was performed by means of the multiply-charged ions from a separate injection of horse heart myoglobin (average molecular mass 16951.5 Da); all masses are reported as average mass.

Acknowledgments

This work was supported in part by CNR grant Progetto Strategico "Biologia Strutturale" and by MURST 40% (Rome).

References

Bianchi E, Sollazzo M, Tramontano A, Pessi A. 1993. Chemical synthesis of a designed β-protein through the flow polyamide method. *Int J Peptide Protein Res 41*:385–393

Bianchi E, Venturini S, Pessi A, Tramontano A, Sollazzo M. 1994. High level expression and rational mutagenesis of a designed protein, the minibody: from an insoluble to a soluble molecule. *J Mol Biol 236*:649–659.

Burnens A, Binz H, Bosshard HR, Corradin G, Demotz S. 1987. Epitope mapping by chemical modification of free and antibody-bound protein antigen. *Science 235*:780–783.

Cuatrecasas P, Fuchs S, Anfinsen CB. 1968. The tyrosyl residues at the active site of staphylococcal nuclease. Modifications by tetranitromethane. *J Biol Chem 243*:4787–4795.

Glazer AN. 1976. The chemical modification of proteins by group-specific and site-specific reagents. In: Neurath H, Hill RL, eds. *The Proteins, 3rd ed., Vol. 2*. New York: Academic Press. pp 1–103.

Glocker MO, Borchers C, Fiedler W, Suckau D, Przybylski M. 1994. Molecular characterization of surface topology in protein tertiary structure by amino-acylation and mass spectrometric peptide. *Bioconj Chem 5*:583–590.

Hsu YR, Narhi LO, Spahr C, Langley KE, Lu HS. 1996. In vitro methionine oxidation of *Escherichia coli*–derived human stem cell factor: Effects on the molecular structure, biological activity, and dimerization. *Protein Sci 5*:1165–1173.

Imoto T, Johnson LN, North ACT, Phillips DC, Rupley JA. 1972. Vertebrate lysozymes. *The Enzymes 7*:666–868.

Loo JA, Edmonds CG, Smith RD. 1990. Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. *Science 248*:201–204.

Lundblad RL, Noyes CN. 1984. *Chemical Reagents for Protein Modification.* Boca Raton, Florida: CRC Press. pp 239–265.

Martin F, Toniatti C, Salvati AL, Ciliberto G, Cortese R, Sollazzo M. 1996. Coupling protein design and in vitro selection strategies: improving specificity and affinity of a designed β-protein IL-6 antagonist. *J. Mol. Biol 255*:86–97.

Martin F, Toniatti C, Salvati AL, Venturini S, Ciliberto G, Cortese R, Sollazzo M. 1994. Affinity selection of a minibody polypeptide inhibitor of human Interleukin-6. *EMBO J 13*:5303–5309.

Ohguro H, Palczewski K, Walsh KA, Johnson RS. 1994. Topographic study of arrestin using differential chemical modifications and hydrogen/deuterium exchange. *Protein Sci 3*:2428–2434.

Pessi A, Bianchi E, Crameri A, Venturini S, Tramontano A, Sollazzo M. 1993. A designed metal binding protein with a novel folding. *Nature 362*:367–369.

Przybylski M, Glocker MO. 1996. Electrospray mass spectrometry of biomacromolecular complexes with noncovalent interactions. New analytical perspectives for supramolecular chemistry and molecular recognition processes. *Angew Chem 35*:806–826.

Sokolovski M, Riordan JF, Vallee BL. 1966. Tetranitromethane. A reagent for the nitration of tyrosyl residues in proteins. *Biochemistry 5*:3582–3589.

Steiner RF, Albaugh S, Fenselau C, Murphy C, Vestling M. 1991. A mass spectrometry method for mapping the interface topography of interacting proteins, illustrated by the Melittin-Calmodulin system. *Anal Biochem 196*:120–125.

Suckau D, Mak M, Przybylski M. 1992. Protein surface topology-probing by selective chemical modification and mass spectrometric peptide mapping. *Proc Natl Acad Sci USA 89*:5630–5634.

Tramontano A, Bianchi E, Venturini S, Martin F, Pessi A, Sollazzo M. 1994. The making of the Minibody: An engineered b-protein for the display of conformationally-constrained peptides. *J Mol Recogn 7*:9–24.

Venturini S, Martin F, Sollazzo M. 1994. Phage display of the Minibody: A β-scaffold for the selection of conformationally-constrained peptides. *Protein Peptide Lett 1*:70–75.

Zappacosta F, Sannia G, Savoy LA, Marino G, Pucci P. 1994. Post-translational modifications in aspartate aminotransferase from *Sulfolobus solfataricus*. Detection of N-ϵ-methyllysines by mass spectrometry. *Eur J Biochem 222*:761–767.

Zappacosta F, Pessi A, Bianchi E, Venturini S, Sollazzo M, Tramontano A, Marino G, Pucci P. 1996. Probing the tertiary structure of proteins by limited proteolysis and mass spectrometry: The case of Minibody. *Protein Sci 5*:802–813.