Protein Science (1995), *4*:2405–2410. Cambridge University Press. Printed in the USA.

Paired natural cysteine mutation mapping: Aid to constraining models of protein tertiary structure

RACHELI KREISBERG, VIRGINIA BUCHNER, AND DORIT ARAD

Department of Molecular Microbiology and Biotechnology, The George S. Wise Faculty of Life Sciences, Tel-Aviv University, Ramat Aviv, Israel 69978

(RECEIVED April 6, 1995; ACCEPTED August 15, 1995)

Abstract

This paper discusses the benefit of mapping paired cysteine mutation patterns as a guide to identifying the positions of protein disulfide bonds. This information can facilitate the computer modeling of protein tertiary structure. First, a simple, paired natural-cysteine-mutation map is presented that identifies the positions of putative disulfide bonds in protein families. The method is based on the observation that if, during the process of evolution, a disulfide-bonded cysteine residue is not conserved, then it is likely that its counterpart will also be mutated. For each target protein, protein databases were searched for the primary amino acid sequences of all known members of distinct protein families. Primary sequence alignment was carried out using PileUp algorithms in the GCG package. To search for correlated mutations, we listed only the positions where cysteine residues were highly conserved and emphasized the mutated residues. In proteins of known three-dimensional structure, a striking pattern of paired cysteine mutations correlated with the positions of known disulfide bridges. For proteins of unknown architecture, the mutation maps showed several positions where disulfide bridging might occur.

Keywords: amino acid sequence; disulfides identification; multiple-sequence alignment; protein folding

Calculating the native conformation of a protein molecule from its amino acid sequence is an intriguing computational problem. Computer simulation of protein folding enables the modeling of protein structure directly from the primary amino acid sequence. Disulfide bonds between cysteine residues within a polypeptide chain produce circular loops that limit the conformations into which a protein can fold. Therefore, when calculating the native-like folded conformation of a protein molecule, correct disulfide bonding is one restraint that may be used to restrict the search for possible conformations. The information for the correct pairing of half-cystine residues in disulfide linkage is encoded in the primary sequence itself (Anfinsen et al., 1961; Anfinsen, 1973). Primary sequence information is now available for more than 15,000 proteins.

Information on disulfide formation can be obtained by time-consuming experimental methods (Hartley et al., 1965; Thannhauser et al., 1984; Morris & Pucci, 1985; Kremser & Rasched, 1994; Xue et al., 1994; Yang et al., 1994). Three decades ago, Hartley et al. (1965) chemically isolated cysteine-bridged peptides from peptide digests of several pancreatic proteolytic enzymes and then compared their amino acid sequences with the primary sequences of their homologs to determine the evolutionary similarities in their disulfide linkages. More recently, Yang et al. (1994) combined Edman degradation and mass spectrometric analysis of a purified cluster of chymotryptic fragments to identify an intramolecular and an intermolecular disulfide bridge in human apolipoprotein D. In certain proteins, however, significant difficulties have been encountered in preparing identifiable or soluble peptides that could permit the explicit identification of cysteines (Xue et al., 1994).

Multiple-sequence alignment between the primary sequence of a protein of unknown structure and one or more homologous protein(s) of known tertiary structure is another approach to assigning disulfide bonds. The main-chain folding for the known protein provides a strong base for structural modeling. This method is limited, however, to cases where the primary structures of the aligned proteins are homologous. Because the primary sequences of most proteins are not homologous to those of any protein of known structure in the Brookhaven Protein Data Bank (PDB) (Bernstein et al., 1977), a method that is independent of the knowledge of the three-dimensional structure is desirable.

Although considerable modification of a protein's amino acid sequence may occur without loss of function, the geometry of the protein, its active site, as well as the residues participating in catalytic or regulatory activities, must be conserved. For example, Altschuh et al. (1988) have shown in homologous protein families that amino acid residues linked by a specific role

Reprint requests to: Dorit Arad, Department of Molecular Microbiology and Biotechnology, The George S. Wise Faculty of Life Sciences, Tel-Aviv University, Ramat Aviv, Israel 69978.

mutate in a coordinated manner if the geometry of the contacts is the same in all related proteins. Horovitz et al. (1994) used this approach to search for possible inter-residue interactions in the *Escherichia coli* chaperonin GroEL, whose three-dimensional structure at atomic resolution remains unknown. By aligning the amino acid sequences of GroEL and its chaperonin 60 protein homologs from other prokaryotic sources, the authors found a significant pattern of correlated mutations in two sites only, which corresponded to positions Cys 137 and Cys 518 in the GroEL protein. A double-mutant cycle procedure was then applied that proved the interaction between these two residues but not to form a disulfide bridge.

For proteins of unknown architecture, knowing which cysteine pairs are disulfide bonded in primary sequences containing more than two cysteines can facilitate the prediction of tertiary structure. Thornton's (1981) extensive and detailed analysis of the conservation of disulfide bridges in globular proteins of known sequence, connectivity, and structure revealed that if a disulfide is not conserved during the process of evolution, then the probability is high that its counterpart will also be replaced. In this report, we attempted to map the pattern of paired natural cysteine mutations at sites corresponding to the positions of conserved cysteine residues in five distinct protein families. For all five families, the natural cysteine mutation map showed one or more coordinated cysteine mutations. The results support the idea that the absence of a highly conserved cysteine pair in homologous proteins is indicative of a cysteine–cysteine interaction in other members of the same family where the same cysteine residues are conserved.

Results and discussion

The number of possible ways (N) to pair an even number of conserved cysteines (n) = {2, 4, 6, . . .} can be calculated by using the following equation (adapted from Kauzmann [1959]):

$$N = \prod_{i=1}^{n} (2i - 1) = \frac{n!}{2^{(n/2)} * (n/2)!}. \quad (1)$$

The possibility of pairing conserved cysteine residues is a function of the number of conserved residues. For example, 2 conserved cysteine residues can be paired in only 1 way to form a disulfide bridge, 4 conserved cysteines can be paired in 3 ways, 6 conserved cysteines can be paired in 15 ways, and 14 cysteines can be paired in 135,135 different ways!

Highly conserved cysteine residues are usually involved in disulfide bridging (Doolittle, 1989). Sequence motifs in diverse proteins of mixed ancestry have been recognized because they contain disulfide-bridged cysteines that are conserved even when most of the surrounding residues have been replaced (Doolittle, 1985). Disulfide bridges are strongly conserved because they contribute to the structure, function, and stability of proteins. In protein superfamilies, however, the conservation rules appear to be less stringent (Thornton, 1981). For example, Figure 1 shows that, in the trypsin-like serine-protease family, trypsin has six disulfide bridges, α-chymotrypsin has five, elastase has four, α-lytic protease has three, and *Streptomyces griseus* proteases A and B have only two.

From analyses of a limited data set, Thornton (1981) has suggested that several guidelines might be used to reduce the number of alternative bridges in proteins for which disulfide

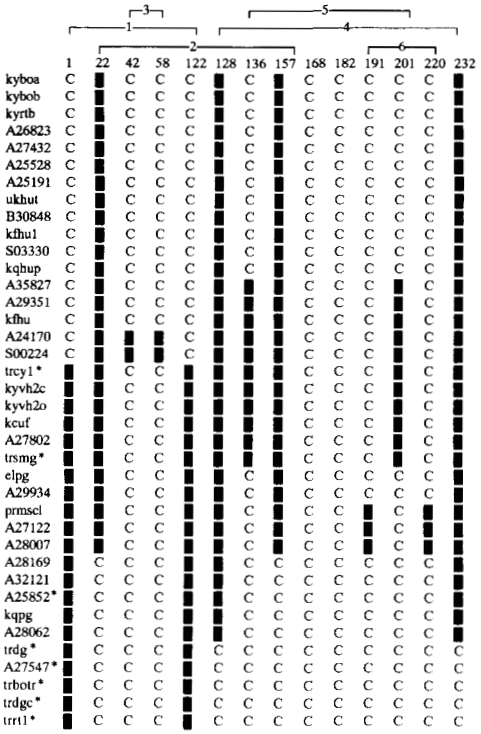

Fig. 1. Conserved and nonconserved cysteine residues in the trypsin-like serine proteinases (from the PDB and Genembl databases). Numbers indicate positions of conserved cysteine residues in the primary sequences. C, conserved cysteine; black rectangle, mutation; *, trypsin proteases. A24170, human complement subcomponent Clr; S00224, human complement subcomponent Cls; A35827, mouse prothrombin; A29351, human thrombin; kfhu, human coagulation factor IXa; kyvh2c, European hornet chymotrypsin II; kyvh2o, Oriental hornet chymotrypsin II; trbotr, bovine trypsin; kcuf, atlantic sand fiddler crab collagenolytic proteinase U; trsmg, *Streptomyces griseus* trypsin; prmscl, mouse cytotoxic T-lymphocyte proteinase 1; A27122, mouse cathepsin G; A28007, rat chymase; kybob, bovine chymotrypsin B; kyrtb, rat chymotrypsin B; kyboa, bovine chymotrypsin A; A26823, pig elastase II; A27432, human elastase II; A25528, mouse elastase II; elpg, pig pancreatic elastase; A29934, human elastase IIIa; A25191, human coagulation factor XIIa; ukhut, human plasminogen activator; B30848, rhesus macaque plasmin; kfhu1, human coagulation factor XIa; kqhup, human plasma kallikrein; trcy1, broadfingered crayfish trypsin; S03330, human acrosin; A27802, early cattle grub hypoderma collagenase; A28169, *Barba amarilla Bothrops atrox* serine proteinase; A32121, Russell's viper proteinase; A25852, human trypsin I; kqpg, pig tissue kallikrein; A28062, mouse γ-renin; trdgc, cationic dog trypsin; trrt1, rat trypsin I; trdg, anionic dog trypsin; A27547, cationic rat trypsin.

connections are not known: (1) "local" (close together in sequence) half-cystines preferentially form disulfide bridges; (2) local or nonlocal disulfides between N/C-termini are also favorable; (3) secondary structure (β-sheet and α-helix) often prevents formation of local disulfides; (4) the loop connecting half-cystines generally includes a tetrapeptide with positive β-turn potential; and (5) disulfides are rarely found between cystines separated by more than 150 residues or between cystines in different domains. Further restrictions occur for local strand-coil and helix-coil connections. How then can one possibly predict from the primary sequence alone which cysteine residues interact?

To address this problem, we used a multiple sequence alignment program to map patterns of correlated cysteine mutations in distinct protein families. Natural cysteine mutation mapping

is based on the observation of Thornton (1981) that when one half-cystine of a disulfide bridge mutates, the remaining thiol group is potentially reactive and mutates rapidly. Thus the number of conserved cysteine residues involved in disulfide bonding should be even. When a cysteine residue has a different function, however, such as participating in a catalytic triad in the papain-like cysteine proteases (Kamphuis et al., 1984), a single conserved cysteine residue exists.

If cysteine residues are mutated at two distinct positions in one or more homologous protein(s) and conserved at the identical positions in all other members of the family, it is likely that the two conserved cysteine residues interact with each other. Such an interaction does not necessarily form a disulfide bridge, however. Even when only two cysteine residues are conserved, automatic pairing to form a disulfide bridge cannot be assumed. For example, Schulteis et al. (1995) have shown that the two conserved cysteine residues in the shaker potassium channel do not form a disulfide bond.

To test the predictive value and validity of mapping natural cysteine mutations, we chose as test cases two enzyme families of known tertiary structure (see the Methods). We aligned the amino acid residues from the family of trypsin-like serine proteinases at 14 positions where cysteine is highly conserved in all 38 proteins (Fig. 1). According to Equation 1, 14 conserved cysteines can be combined in 135,135 ways. A glance at Figure 1 reveals a striking pattern of 12 paired cysteine mutations (solid bars) at the following positions: 1–122, 22–157, 42–58, 128–232, 136–201, and 191–220.

It should be noted that Cys 122 is actually unpaired in the mature chymotrypsin molecule. In some precursors of the trypsin-like serine proteinases, however, Cys 122 is paired with a conserved cysteine at position 1. Although natural mutations were absent in positions 168 and 182, we deduced that this pair also forms a seventh disulfide bridge. Examination of the known three-dimensional structures of trypsin, α-trypsin, chymotrypsin A, elastase, and kallikrein A confirmed the location of disulfide bridges at these 14 positions.

When all cysteine residues appear to be conserved at certain positions in a family of proteins, sequences of lower homology must be included in the natural cysteine mutation map to successfully identify disulfide interactions. Hence, the mutation map should include the primary sequences of all known members of a distinct family to increase the possibility that at least one mutant cysteine pair will be found. For example, in 14 of the 20 members of the trypsin family, the disulfide bridges are strictly conserved (data not shown). If the trypsin proteases (marked with * in Fig. 1) had been mapped without including all the trypsin-like serine proteinases, the disulfide bridges at positions Cys 42 and Cys 58 (mutated in human complement subcomponents C1r and C1s) and positions Cys 191 and Cys 220 (mutated in mouse cytotoxic T-lymphocyte proteinase I, rat chymase, and mouse cathepsin G) would not be detectable, and the 168–182 and 191–200 disulfide links could not have been deduced because they are close in the sequence. Thus, only three of the disulfide bridges (22–157, 128–232, 136–201) in the known tertiary structure would have been detected because the other disulfide-bridged cysteines are strictly conserved.

In certain protein families, the paired natural cysteine-mutation mapping method is limited by the absence of mutant proteins. It is possible that all homologous primary sequences in a given family have not yet been determined. Phospholipase A2 is one example where most cysteine residues are conserved in all known family members. Analyzing the data in Figure 2 reveals that cysteine residues are mutated at positions 11 and 77 in seven family members. The conserved cysteines at positions 50 and 133 in five of these proteins have been replaced in all other members of this family. Two of these proteins are also the only family members that show mutations at positions 61 and 91. The cysteines at all other positions are fully conserved in all proteins from this family. Unless primary sequences from other family members are deduced, our mutation map can detect only three out of eight disulfide bridges: Cys 11–Cys 77, Cys 50–Cys 133, and Cys 61–Cys 91. We verified the location of these bridges in three tertiary structures obtained from the PDB. Although the 16 highly conserved cysteine residues in the phospholipase A2 family can be combined in 2,027,025 ways according to Equation 1, the successful prediction of three of eight disulfide-bridge locations excludes 2,026,080 of the possible combinations. The field of search is thus reduced considerably because only 945 ways remain to pair the remaining 10 residues.

When three distinct cysteine positions showed identical mutation patterns, we deduced possible disulfide interactions according to Thornton's (1981) guidelines. The same general guidelines should also apply in the case of more than two completely conserved cysteine positions, such as in phospholipase A2, where the strictly conserved Cys 27 and Cys 123 residues have been shown to link the N- and C-termini (Thornton, 1981).

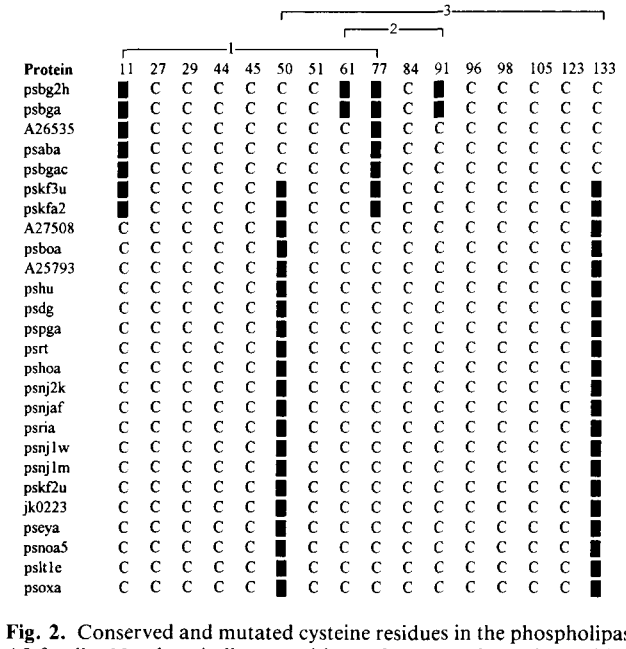

Fig. 2. Conserved and mutated cysteine residues in the phospholipase A2 family. Numbers indicate positions of conserved cysteine residues in the primary sequences of phospholipase A2. C, conserved cysteine; black rectangle, mutation. psbg2h, rhinoceros viper; psbga, gaboon viper; A26535, halys viper; psaba, mamushi; psbgac, horned viper; pskf3u, β-bungarotoxin A2 chain in many-banded krait; pskfa2, β-bungarotoxin A2 chain in many-banded krait; A27508, bovine; psboa, bovine; A25793, human; pshu, human; psdg, dog; pspga, pig; psrt, rat; pshoa, horse; psnj2k, monocled cobra; psnjaf, Chinese cobra; psria, ringhals venom; psnj1w, forest cobra; psnj1m, mozambique cobra; pskf2u, many-banded krait; jk0223, banded krait venom; pseya, beaked sea snake; psnoa5, common tiger snake; pslt1e, broad-banded blue sea krait; psoxa, Australian taipan.

We attempted to map paired cysteine mutations in two subfamilies of the protein kinases for whom the tertiary structures are unknown (Figs. 3, 4). For the protein kinases listed in Figure 3, we would predict interacting cysteine residues at positions 466–549, 621–705, 757–765, 891–1014, 1136–1147, and 1216–1236. In Figure 4, the problem is more complicated because three possibilities exist for interactions of positions Cys 65, Cys 150, and Cys 211, which show identical mutant patterns. Moreover, both Cys 65 and Cys 211 are less than 150 amino acids from position Cys 150. Nevertheless, cysteines at positions 309–372, 652–727, and 1107–1271 can be predicted to interact. Because the three-dimensional structures of proteins from these families have not yet been determined, the possibility of cysteine–cysteine interactions at these positions must be confirmed experimentally. Figure 5 shows the mutation map of the γ-glutamyl transferase family. In this situation, we inferred that Cys 32 and Cys 490 may form a disulfide bridge to join the N/C-termini.

The question arises of whether identifying the locations of disulfides can facilitate the prediction of the tertiary structure of a protein. The disappearance or evolution of disulfide bridges in related proteins, as seen in our mutation maps, appears to contradict the conventional wisdom that disulfide bridges are important to the protein-folding process. Crystallographic data imply that the absence of a disulfide bridge in one family member does not always change the basic structure of that protein. For example, the two disulfide bonds found only in trypsin (Cys 22–Cys 157 and Cys 128–Cys 232 in Fig. 1) fit easily into the models of elastase and α-chymotrypsin, almost without altering their three-dimensional structures (Thornton, 1981).

Many investigators have found that the absence of one or more disulfides does not change the tertiary structure of the protein under study (Dai & Tang, 1994; Eyles et al., 1994; Kumar et al., 1994; Sery et al., 1994). Dai and Tang (1994), for example, have shown that although the intra-A chain disulfide bond (A6–11) of insulin is essential for activity, the three-dimensional structure was not altered significantly by its absence. This report is interesting because multiple sequence alignment of 39 insulin family members showed that all six cysteine residues are

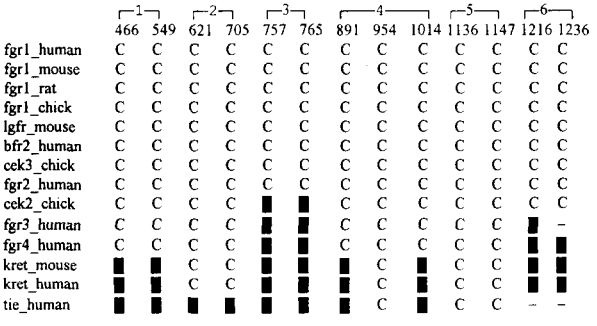

Fig. 3. Conserved and mutated cysteine residues in the tyrosine kinases subfamily. Numbers indicate positions of conserved cysteine residues in the primary sequences. C, conserved cysteine; black rectangle, mutation. fgr1, basic fibroblast growth factor receptor 1 precursor; fgr3, fibroblast growth factor receptor 3 precursor; fgr4, fibroblast growth factor receptor 4 precursor; bgr4, fibroblast growth factor receptor bfr-2 precursor; kgfr, keratinocyte, growth factor receptor precursor; cek2, tyrosine kinase receptor cek2 precursor; kret, proto-oncogene tyrosine-protein kinase receptor ret precursor; tie, tyrosine-protein kinase receptor tie precursor.

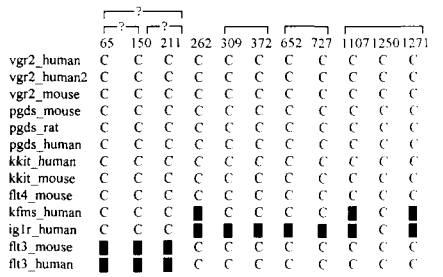

Fig. 4. Conserved and mutated cysteine residues in the tyrosine protein kinases. Numbers indicate positions of conserved cysteine residues in the primary sequences. C, conserved cysteine; black rectangle, mutation. vgr2, vascular endothelial growth factor receptor 2 precursor; pgds, α-platelet-derived growth factor receptor precursor; kkit, mast/stem cell growth factor receptor precursor; flt4, tyrosine-protein kinase receptor flt4 precursor; kfms, macrophage colony stimulating factor i receptor precursor; iglr, insulin-like growth factor I receptor precursor; flt3, fl cytokine receptor precursor.

fully conserved (data not shown). Kumar et al. (1994) presented an analysis of the soluble species formed on refolding of RNase A that appears to reaffirm the principle that the information for folding of the protein is encoded in the amino acid sequence itself.

Human chorionic gonadotropin (hCG) is a member of a family of heterodimeric glycoprotein hormones that contain a common α subunit but differ in their hormone-specific β subunits. In the hCG α subunit, five cysteines participating in disulfide bond formation are not essential for the α subunit to fold into a form that will combine with the β subunit and to produce a biologically active dimer (Furuhashi et al., 1994). In the hCG β subunit, however, all the disulfide bonds were found to be required for efficient combination and folding (Suganuma et al., 1989). Bedows et al. (1993) presented evidence to show that hCG-β folding coincides with the formation of specific disulfide bonds. The folding behavior of the hCG β subunit supports the hypothesis that individual disulfide bonds are involved in discrete steps in the hCG β folding pathway so that the final tertiary structure may be achieved only after disulfide bridging has occurred during the early steps of the folding process. Goldberg and Guillou (1994) reached a similar conclusion while studying secondary structure formation during the folding of lysozyme.

This notion is supported further by the results of recent research on native and recombinant insulin-like growth factor (IGF-I) (Axelsson et al., 1992; Narhi et al., 1993; Miller et al.,

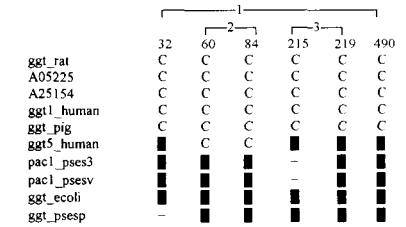

Fig. 5. Conserved and mutated cysteine residues in the γ-glutamyl transferases. Numbers indicate positions of conserved cysteine residues in the primary sequences. C, conserved cysteine; black rectangle, mutation; –, gap in sequence.

1993; Yan & Erickson, 1994), which in aqueous solution folds into a spatial structure similar to that of insulin (Sato et al., 1993). Axelsson et al. (1992) reported that correct disulfide arrangement is important for the full biological activity of IGF-I. Subsequently, Narhi et al. (1993) found that the native structure, function, and stability of IGF-I requires the presence of all three disulfide bonds. The results of the oxidative refolding experiments of Miller et al. (1993) suggest that one IGF-1 protein sequence may encode two distinct tertiary folds. In the chemically synthesized, disulfide-bridged, covalent dimer β-bellin 14D, Yan and Erickson (1994) found that folding was induced by formation of the single interchain disulfide bond, but not by the binary pattern of alternating polar and nonpolar residues of its β-sheets. Disulfide-bond formation has been implicated in both the folding and the tetramerization of a recombinant human thyrotropin (TSH) receptor as well (Graves et al., 1995).

Thus, we believe that disulfide-bridge measurements are important for constraining the calculations of protein tertiary structures. We have shown here that mapping the patterns of natural cysteine mutations can be used to predict such cysteine–cysteine interactions whenever such information is needed for this task. Calculating protein folding by restrained energy minimization is one example where correct disulfide bonding has been used as a major restraint to restrict the search for conformations having the lowest possible values of free and potential energy (Levitt, 1983). In another example, Sato et al. (1993) included 12 interatomic distance constraints related to disulfide bridges in the distance geometry calculations that were combined with NMR measurements to determine the three-dimensional structure of human IGF-I. Disulfide bridge locations predicted by natural cysteine mutation mapping may be combined with any other constraint that is used for secondary structure prediction, turn predictions, molecular dynamics simulation, and protein modeling.

The number of known tertiary structures is still relatively small compared with the number of known primary sequences, which underscores the need for a method that can derive structural features solely from the primary sequence. Present methods for determining the location of disulfide bridges require either experimental results or the knowledge of homologous three-dimensional structures. Natural cysteine mutation mapping, on the other hand, depends only on multiple primary sequence alignment in protein families with homologous primary amino acid sequences. We do not know at the present time, however, exactly how many families of extracellular protein exist for which the three-dimensional structure is not yet known. This method combines the knowledge that has already accumulated in large experimental databases with user-friendly computational software. Despite its inherent limitations, we believe that natural cysteine mutation mapping is a rapid, useful, and convenient guide to identifying putative disulfide bridges. Such information can be useful for constraining computer models of protein tertiary structures even if it suggests only part of the pattern.

Methods

To map natural cysteine mutations in target proteins, we identified the primary sequences of distinct protein families. We searched the National Biomedical Research Foundation Protein Sequence Database (George et al., 1986), and the SwissProt Protein Sequence Database, and the GenBank® Nucleic Acid Sequence Database (Burks et al., 1985) for the primary sequences of all known family members, using the FASTA and TFASTA programs (Lipman & Pearson, 1985) of GCG (Devereux et al., 1984). Multiple sequence alignment of the maximum number of sequences belonging to a family was performed with the PileUp algorithms in the GCG package (Needleman, & Wunsch, 1970; Devereux et al., 1984).

The alignment "score" for a given residue refers to its priority to be properly aligned. The strictly conserved catalytic amino acids of enzyme families have the highest score and thus are automatically aligned in the first step. If a catalytic amino acid was slightly misaligned, we changed the algorithm to raise the score of the misaligned catalytic residue to bring it into alignment with the other catalytic residues. We used the second-highest score to align the cysteine residues. If a cysteine residue had become misaligned as a result of adjusting the positions of the catalytic residues, the score of the cysteine residue was manually changed to bring it into alignment with the other cysteines.

For each target protein, protein databases were searched for the primary amino acid sequences of all known members of distinct protein families. To search for correlated mutations, we listed only the positions where cysteine residues were highly conserved. The mutated, noncysteine amino acids were then marked with solid bars to emphasize their positions, and the protein list was rearranged so that the mutated residues would be contiguous. We assigned putative disulfide interactions between cysteine residues when identical mutation patterns appeared at two positions that were occupied by conserved cysteines in all other members of the family.

Acknowledgments

We thank Dr. H. Prilusky of the Weizmann Institute of Science, Rehovot, Israel, for his outstanding help, as well as Dr. Amnon Horovitz for useful discussions of his work. We also thank Pinchas Shamgar from our group for performing the database search, multiple sequence alignment, and PileUp for the tyrosine protein kinase families.

References

Altschuh D, Vernet T, Berti P, Moras D, Nagai K. 1988. Coordinated amino acid changes in homologous protein families. *Protein Eng 2*:193–199.

Anfinsen CB. 1973. Principles that govern the folding of protein chains. *Science 181*:223–230.

Anfinsen CB, Haber E, Sela M, White FH Jr. 1961. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci USA 47*:1309–1314.

Axelsson K, Johansson S, Eketorp G, Zazzi H, Hemmendorf B, Gellerfors P. 1992. Disulfide arrangement of human insulin-like growth factor I derived from yeast and plasma. *Eur J Biochem 206*:987–994.

Bedows E, Huth JR, Suganuma N, Bartels CF, Boime I, Ruddon RW. 1993. Disulfide bond mutations affect the folding of the human chorionic gonadotropin-β subunit in transfected Chinese hamster ovary cells. *J Biol Chem 268*:11655–11662.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rogers 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.

Burks C, Fickett JW, Goad WB, Kanehisa M, Lewitter FI, Rindone WP, Swindell CD, Tung CS, Bilofsky HS. 1985. The GenBank nucleic acid sequence database. *Comput Appl Biosci 1*:225–233.

Dai Y, Tang JG. 1994. Intra-A chain disulfide bond (A6–11) of insulin is essential for displaying its activity. *Biochem Mol Biol Int 33*:1049–1053.

Devereux J, Haeberli P, Smithies O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res 12*:387–395.

Dickerson RE, Timkovich R. Cytochromes C. 1975. In: Boyer PD, ed. *The enzymes, vol 11*. New York: Academic Press. pp 397–593.

Doolittle RF. 1985. The genealogy of some recently evolved vertebrate proteins. *Trends Biochem Sci 10*:233–237.

Doolittle RF. 1989. Similar amino acid sequences revisited. *Trends Biochem Sci 14*:244–245.

Eyles SJ, Radford SE, Robinson CV, Dobson CM. 1994. Kinetic consequences of the removal of a disulfide bridge on the folding of hen lysozyme. *Biochemistry 33*:13038.

Furuhashi M, Ando H, Bielinska M, Pixley MR, Shikone T, Hsueh AJ, Boime I. 1994. Mutagenesis of cysteine residues in the human gonadotropin alpha subunit. Roles of individual disulfide bonds in secretion, assembly, and biologic activity. *J Biol Chem 269*:25543–25548

George DG, Barker WC, Hunt LT. 1986. The protein identification resource PIR. *Nucleic Acids Res 14*:11–15.

Goldberg ME, Guillou AD. 1994. Native disulfide bonds greatly accelerate secondary structure formation in the folding of lysozyme. *Protein Sci 3*:883–887.

Graves PN, Vlase H, Davies TF. 1995. Folding of the recombinant human thyrotropin (TSH) receptor extracellular domain: Identification of folded monomeric and tetrameric complexes that bind TSH receptor autoantibodies. *Endocrinology 136*:521–527.

Hartley BS, Brown JR, Kauffman DL, Smillie LB. 1965. Evolutionary similarities between pancreatic proteolytic enzymes. *Nature 207*:1157–1159.

Horovitz A, Bochkareva ES, Yifrach O, Girshovich AS. 1994. Prediction of an inter-residue interaction in the chaperonin GroEL from multiple sequence alignment is confirmed by double-mutant cycle analysis. *J Mol Biol 238*:133–138.

Kamphuis IG, Kalk KH, Swarte MBA, Drenth J. 1984. Structure of papain refined at 165 angstroms resolution. *J Mol Biol 179*:233–256.

Kauzmann W. 1959. Relative probabilities of isomers in cystine-containing randomly coiled polypeptides. In: Benesch GA, Boyer P, Klotz I, Middlebrook WR, Szent-Gyorgyi A, Schwarz DR, eds. *Sulfur in proteins*. New York: Academic Press. pp 93–108.

Kremser A, Rasched I. 1994. The adsorption protein of filamentous phage fd: Assignment of its disulfide bridges and identification of the domain incorporated in the coat. *Biochemistry 33*:13954–13958.

Kumar TK, Gopalakrishna K, Ramakrishna T, Pandit MW. 1994. Refolding of RNAse A at high concentrations: Identification of non-native species. *Int J Biol Macromol 16*:171–176.

Kuntz ID, Crippen GM, Kollman PA, Kimelman D. 1976. Calculation of protein tertiary structure. *J Mol Biol 106*:983–994.

Levitt M. 1983. Protein folding by restrained energy minimization and molecular dynamics. *J Mol Biol 170*:723–764.

Lipman DJ, Pearson WR. 1985. Rapid and sensitive protein similarity searches. *Science 227*:1435–1441.

Miller JA, Narhi LO, Hua QX, Rosenfeld R, Arakawa T, Rohde M, Prestrelski S, Lauren S, Stoney KS, Tsai L, et al. 1993. Oxidative refolding of insulin-like growth factor I yields two products of similar thermodynamic stability: A bifurcating protein-folding pathway. *Biochemistry 32*:5203–5213.

Morris HR, Pucci P. 1985. A new method for rapid assignment of S–S bridges in proteins. *Biochem Biophys Res Commun 126*:1122–1128.

Narhi LO, Hua QX, Arakawa T, Fox GM, Rtsai L, Rosenfeld R, Holst P, Miller JA, Weiss MA. 1993. Role of native disulfide bonds in the structure and activity of insulin-like growth factor 1: Genetic models of protein-folding intermediates. *Biochemistry 32*:5214–5221.

Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol 48*:443–453.

Sato A, Nishimura S, Ohkubo T, Kyogoku Y, Koyama S, Kobayashi M, Yasuda T, Kobayashi Y. 1993. Three-dimensional structure of human insulin-like growth factor-I(IGF-I) determined by ^{1}H-NMR and distance geometry. *Int J Pept Protein Res 421*:433–440.

Schulteis CT, John SA, Huang Y, Tang CY, Papazian DM. 1995. Conserved cysteine residues in the shaker K^+ channel are not linked by a disulfide bond. *Biochemistry 34*:1725–17233.

Sery A, Housset D, Serre L, Bonicel J, Hatchikian C, Frey M, Roth M. 1994. Crystal structure of the ferredoxin I from *Desulfovibrio africanus* at 2.3 Angstrom resolution. *Biochemistry 33*:15408–15417.

Suganuma N, Matzuk M, Boime I. 1989. Elimination of disulfide bonds affects assembly and secretion of the human chorionic gonadotropin beta subunit. *J Biol Chem 264*:19302–19307.

Thannhauser TW, Konishi Y, Scheraga HA. 1984. Sensitive quantitative analysis of disulfide bonds in polypeptides and proteins. *Anal Biochem 138*:181–188.

Thornton JM. 1981. Disulphide bridges in globular proteins. *J Mol Biol 151*:261–187.

Xue J, Kalafatis M, Silveira JR, Kung C, Mann KG. 1994. Determination of the disulfide bridges in factor Va heavy chain. *Biochemistry 33*:13109–13116.

Yan Y, Erickson BW. 1994. Engineering of betabellin 14D: Disulfide-induced folding of a beta-sheet protein. *Protein Sci 3*:1069–1073.

Yang CY, Gu ZW, Blanco-Vaca F, Gaskell SJ, Yang M, Massey JB, Gotto AM Jr, Pownall HJ. 1994. Structure of human apolipoprotein D: Locations of the intermolecular and intramolecular disulfide links. *Biochemistry 33*:12451–12455.