Finally, CSD data are used in several *de novo* design programs. These types of programs, *e.g. LUDI* (Böhm, 1992*a*,*b*), predict novel ligands that will interact favourably with a given protein and use hydrogen-bond geometries from the CSD (indirectly) to position their structural fragments in the binding site.

22.4.6. Conclusion

This chapter has summarized the vast range of structural knowledge that can be derived from the small-molecule data contained in the CSD. We have attempted to show that much of this knowledge is directly transferable and applicable to the protein environment. Far from being discrete, structural studies of small molecules and proteins have a natural synergy which, if exploited creatively, will lead to significant advances in both areas. It is therefore unsurprising that some of these CSD studies have been prompted by initial observations made on proteins.

As a result of this activity, it is now very clear that software access to the information stored in the CSD and the PDB must be at two levels: a raw-data level and a derived-knowledge level. The onward development of structural knowledge bases from the underlying data provides for the preservation and storage of the results of data-mining experiments, thus avoiding repetition of standard experiments and providing instant access to complex derivative information. Most importantly, a suitably structured knowledge base can be acted on by software tools that are designed to solve complex problems in structural chemistry (see *e.g.* Thornton & Gardner, 1989; Allen *et al.*, 1990; Bruno *et al.*, 1997; Jones *et al.*, 1997). The availability of knowledge bases derived from experimental observations is likely to be a crucial factor in the solution of those two analogous, and currently intractable, problems in the small-molecule and protein-structure domains: crystal structure and polymorph prediction on the one hand, and protein folding on the other.

References

22.1

Acharya, R., Fry, E., Logan, D., Stuart, D., Brown, F., Fox, G. & Rowlands, D. (1990). *The three-dimensional structure of foot-and-mouth disease virus. New aspects of positive-strand RNA viruses*, edited by M. A. Brinton & S. X. Heinz, pp. 319–327. Washington DC: American Society for Microbiology.

Arnold, E. & Rossmann, M. G. (1990). *Analysis of the structure of a common cold virus, human rhinovirus 14, refined at a resolution of 3.0 Å. J. Mol. Biol.* **211**, 763–801.

Baker, E. N. & Hubbard, R. E. (1984). *Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol.* **44**, 97–179.

Bernal, J. D. & Finney, J. L. (1967). *Random close-packed hard-sphere model II. Geometry of random packing of hard spheres. Discuss. Faraday Soc.* **43**, 62–69.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1965). *Structure of hen egg-white lysozyme, a three-dimensional Fourier synthesis at 2 Å resolution. Nature (London)*, **206**, 757–761.

Bondi, A. (1964). *van der Waals volumes and radii. J. Phys. Chem.* **68**, 441–451.

Bondi, A. (1968). *Molecular crystals, liquids and glasses*. New York: Wiley.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). *CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem.* **4**, 187–217.

Chandler, D., Weeks, J. D. & Andersen, H. C. (1983). *van der Waals picture of liquids, solids, and phase transformations. Science*, **220**, 787–794.

Chapman, M. S. (1993). *Mapping the surface properties of macromolecules. Protein Sci.* **2**, 459–469.

Chapman, M. S. (1994). *Sequence similarity scores and the inference of structure/function relationships. Comput. Appl. Biosci. (CABIOS)*, **10**, 111–119.

Chothia, C. (1975). *Structural invariants in protein folding. Nature (London)*, **254**, 304–308.

Chothia, C. (1976). *The nature of the accessible and buried surfaces in proteins. J. Mol. Biol.* **105**, 1–12.

Chothia, C. & Janin, J. (1975). *Principles of protein–protein recognition. Nature (London)*, **256**, 705–708.

Connolly, M. (1986). *Measurement of protein surface shape by solid angles. J. Mol. Graphics*, **4**, 3–6.

Connolly, M. L. (1983). *Analytical molecular surface calculation. J. Appl. Cryst.* **16**, 548–558.

Connolly, M. L. (1991). *Molecular interstitial skeleton. Comput. Chem.* **15**, 37–45.

Diamond, R. (1974). *Real-space refinement of the structure of hen egg-white lysozyme. J. Mol. Biol.* **82**, 371–391.

Dunfield, L. G., Burgess, A. W. & Scheraga, H. A. (1979). *J. Phys. Chem.* **82**, 2609.

Edelsbrunner, H., Facello, M. & Liang, J. (1996). *On the definition and construction of pockets in macromolecules*, pp. 272–287. Singapore: World Scientific.

Edelsbrunner, H., Facello, M., Ping, F. & Jie, L. (1995). *Measuring proteins and voids in proteins. Proc. 28th Hawaii Intl Conf. Sys. Sci.* pp. 256–264.

Edelsbrunner, H. & Mucke, E. (1994). *Three-dimensional alpha shapes. ACM Trans. Graphics*, **13**, 43–72.

Eisenberg, D. & McLachlan, A. D. (1986). *Solvation energy in protein folding and binding. Nature (London)*, **319**, 199–203.

Fauchere, J.-L. & Pliska, V. (1983). *Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. Eur. J. Med. Chem. Chim. Ther.* **18**, 369–375.

Finkelstein, A. (1994). *Implications of the random characteristics of protein sequences for their three-dimensional structure. Curr. Opin. Struct. Biol.* **4**, 422–428.

Finney, J. L. (1975). *Volume occupation, environment and accessibility in proteins. The problem of the protein surface. J. Mol. Biol.* **96**, 721–732.

Finney, J. L., Gellatly, B. J., Golton, I. C. & Goodfellow, J. (1980). *Solvent effects and polar interactions in the structural stability and dynamics of globular proteins. Biophys. J.* **32**, 17–33.

Fritz-Wolf, K., Schnyder, T., Wallimann, T. & Kabsch, W. (1996). *Structure of mitochondrial creatine kinase. Nature (London)*, **381**, 341–345.

Gelin, B. R. & Karplus, M. (1979). *Side-chain torsional potentials: effect of dipeptide, protein, and solvent environment. Biochemistry*, **18**, 1256–1268.

Gellatly, B. J. & Finney, J. L. (1982). *Calculation of protein volumes: an alternative to the Voronoi procedure. J. Mol. Biol.* **161**, 305–322.

Gerstein, M. (1992). *A resolution-sensitive procedure for comparing surfaces and its application to the comparison of antigen-combining sites. Acta Cryst.* A**48**, 271–276.

Gerstein, M. & Chothia, C. (1996). *Packing at the protein–water interface. Proc. Natl Acad. Sci. USA*, **93**, 10167–10172.

Gerstein, M., Lesk, A. M., Baker, E. N., Anderson, B., Norris, G. & Chothia, C. (1993). *Domain closure in lactoferrin: two hinges produce a see-saw motion between alternative close-packed interfaces. J. Mol. Biol.* **234**, 357–372.

Gerstein, M., Lesk, A. M. & Chothia, C. (1994). *Structural mechanisms for domain movements. Biochemistry*, **33**, 6739–6749.

Gerstein, M. & Lynden-Bell, R. M. (1993*a*). *Simulation of water around a model protein helix. 1. Two-dimensional projections of solvent structure. J. Phys. Chem.* **97**, 2982–2991.

22.1 (*cont.*)

Gerstein, M. & Lynden-Bell, R. M. (1993*b*). *Simulation of water around a model protein helix. 2. The relative contributions of packing, hydrophobicity, and hydrogen bonding*. *J. Phys. Chem.* **97**, 2991–2999.

Gerstein, M. & Lynden-Bell, R. M. (1993*c*). *What is the natural boundary for a protein in solution? J. Mol. Biol.* **230**, 641–650.

Gerstein, M., Sonnhammer, E. & Chothia, C. (1994). *Volume changes on protein evolution*. *J. Mol. Biol.* **236**, 1067–1078.

Gerstein, M., Tsai, J. & Levitt, M. (1995). *The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra*. *J. Mol. Biol.* **249**, 955–966.

Grant, J. A. & Pickup, B. T. (1995). *A Gaussian description of molecular shape. J. Phys. Chem.* **99**, 3503–3510.

Greer, J. & Bush, B. L. (1978). *Macromolecular shape and surface maps by solvent exclusion. Proc. Natl Acad. Sci. USA*, **75**, 303–307.

Harber, J., Bernhardt, G., Lu, H.-H., Sgro, J.-Y. & Wimmer, E. (1995). *Canyon rim residues, including antigenic determinants, modulate serotype-specific binding of polioviruses to mutants of the poliovirus receptor. Virology*, **214**, 559–570.

Harpaz, Y., Gerstein, M. & Chothia, C. (1994). *Volume changes on protein folding*. *Structure*, **2**, 641–649.

Hermann, R. B. (1977). *Use of solvent cavity area and number of packed solvent molecules around a solute in regard to hydrocarbon solubilities and hydrophobic interactions*. *Proc. Natl Acad. Sci. USA*, **74**, 4144–4195.

Hubbard, S. J. & Argos, P. (1994). *Cavities and packing at protein interfaces*. *Protein Sci.* **3**, 2194–2206.

Hubbard, S. J. & Argos, P. (1995). *Evidence on close packing and cavities in proteins*. *Curr. Opin. Biotechnol.* **6**, 375–381.

Kapp, O. H., Moens, L., Vanfleteren, J., Trotman, C. N. A., Suzuki, T. & Vinogradov, S. N. (1995). *Alignment of 700 globin sequences: extent of amino acid substitution and its correlation with variation in volume*. *Protein Sci.* **4**, 2179–2190.

Kauzmann, W. (1959). *Some factors in the interpretation of protein denaturation. Adv. Protein Chem.* **14**, 1–63.

Kelly, J. A., Sielecki, A. R., Sykes, B. D., James, M. N. & Phillips, D. C. (1979). *X-ray crystallography of the binding of the bacterial cell wall trisaccharaide NAM-NAG-NAM to lysozymes. Nature (London)*, **282**, 875–878.

Kim, K. H., Willingmann, P., Gong, Z. X., Kremer, M. J., Chapman, M. S., Minor, I., Oliveira, M. A., Rossmann, M. G., Andries, K., Diana, G. D., Dutko, F. J., McKinlay, M. A. & Pevear, D. C. (1993). *A comparison of the anti-rhinoviral drug binding pocket in HRV14 and HRV1A. J. Mol. Biol.* **230**, 206–227.

Kleywegt, G. J. & Jones, T. A. (1994). *Detection, delineation, measurement and display of cavities in macromolecular structures*. *Acta Cryst.* D**50**, 178–185.

Kocher, J. P., Prevost, M., Wodak, S. J. & Lee, B. (1996). *Properties of the protein matrix revealed by the free energy of cavity formation*. *Structure*, **4**, 1517–1529.

Kraulis, P. J. (1991). *MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst.* **24**, 946–950.

Kuhn, L. A., Siani, M. A., Pique, M. E., Fisher, C. L., Getzoff, E. D. & Tainer, J. A. (1992). *The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures*. *J. Mol. Biol.* **228**, 13–22.

Lee, B. & Richards, F. M. (1971). *The interpretation of protein structures: estimation of static accessibility*. *J. Mol. Biol.* **55**, 379–400.

Leicester, S. E., Finney, J. L. & Bywater, R. P. (1988). *Description of molecular surface shape using Fourier descriptors*. *J. Mol. Graphics*, **6**, 104–108.

Levitt, M., Hirshberg, M., Sharon, R. & Daggett, V. (1995). *Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution*. *Comput. Phys. Comm.* **91**, 215–231.

Lewis, M. & Rees, D. C. (1985). *Fractal surfaces of proteins*. *Science*, **230**, 1163–1165.

Lim, V. I. & Ptitsyn, O. B. (1970). *On the constancy of the hydrophobic nucleus volume in molecules of myoglobins and hemoglobins*. *Mol. Biol. (USSR)*, **4**, 372–382.

Madan, B. & Lee, B. (1994). *Role of hydrogen bonds in hydrophobicity: the free energy of cavity formation in water models with and without the hydrogen bonds*. *Biophys. Chem.* **51**, 279–289.

Matthews, B. W., Morton, A. G. & Dahlquist, F. W. (1995). *Use of NMR to detect water within nonpolar protein cavities*. (Letter.) *Science*, **270**, 1847–1849.

Merritt, E. A. & Bacon, D. J. (1997). *Raster3D: photorealistic molecular graphics. Methods Enzymol.* **277**, 505–525.

Molecular Structure Corporation (1995). *Insight II user guide*. Biosym/MSI, San Diego.

Nemethy, G., Pottle, M. S. & Scheraga, H. A. (1983). *Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions and hydrogen bond interactions for the naturally occurring amino acids*. *J. Phys. Chem.* **87**, 1883–1887.

Nicholls, A. (1992). *GRASP: graphical representation and analysis of surface properties.* New York: Columbia University.

Nicholls, A. & Honig, B. (1991). *A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. J. Comput. Chem.* **12**, 435–445.

Nicholls, A., Sharp, K. & Honig, B. (1991). *Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins*, **11**, 281–296.

Olson, N., Kolatkar, P., Oliveira, M. A., Cheng, R. H., Greve, J. M., McClelland, A., Baker, T. S. & Rossmann, M. G. (1993). *Structure of a human rhinovirus complexed with its receptor molecule. Proc. Natl Acad. Sci. USA*, **90**, 507–511.

O'Rourke, J. (1994). *Computational geometry in C*. Cambridge University Press.

Palmenberg, A. C. (1989). *Sequence alignments of picornaviral capsid proteins*. In *Molecular aspects of picornavirus infection and detection*, edited by B. L. Semler & E. Ehrenfeld, pp. 211–241. Washington DC: American Society for Microbiology.

Pattabiraman, N., Ward, K. B. & Fleming, P. J. (1995). *Occluded molecular surface: analysis of protein packing*. *J. Mol. Recognit.* **8**, 334–344.

Pauling, L. (1960). *The nature of the chemical bond*, 3rd ed. Ithaca: Cornell University Press.

Peters, K. P., Fauck, J. & Frommel, C. (1996). *The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria*. *J. Mol. Biol.* **256**, 201–213.

Petitjean, M. (1994). *On the analytical calculation of van der Waals surfaces and volumes: some numerical aspects*. *J. Comput. Chem.* **15**, 1–10.

Pontius, J., Richelle, J. & Wodak, S. J. (1996). *Deviations from standard atomic volumes as a quality measure for protein crystal structures*. *J. Mol. Biol.* **264**, 121–136.

Procacci, P. & Scateni, R. (1992). *A general algorithm for computing Voronoi volumes: application to the hydrated crystal of myoglobin*. *Int. J. Quant. Chem.* **42**, 151–152.

Rashin, A. A., Iofin, M. & Honig, B. (1986). *Internal cavities and buried waters in globular proteins*. *Biochemistry*, **25**, 3619–3625.

Reynolds, J. A., Gilbert, D. B. & Tanford, C. (1974). *Empirical correlation between hydrophobic free energy and aqueous cavity surface area*. *Proc. Natl Acad. Sci. USA*, **71**, 2925–2927.

Richards, F. M. (1974). *The interpretation of protein structures: total volume, group volume distributions and packing density*. *J. Mol. Biol.* **82**, 1–14.

Richards, F. M. (1977). *Areas, volumes, packing, and protein structure*. *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.

Richards, F. M. (1979). *Packing defects, cavities, volume fluctuations, and access to the interior of proteins. Including some general comments on surface area and protein structure*. *Carlsberg Res. Commun.* **44**, 47–63.

22.1 (*cont.*)

Richards, F. M. (1985). *Calculation of molecular volumes and areas for structures of known geometry*. *Methods Enzymol.* **115**, 440–464.

Richards, F. M. & Lim, W. A. (1994). *An analysis of packing in the protein folding problem*. *Q. Rev. Biophys.* **26**, 423–498.

Richmond, T. J. (1984). *Solvent accessible surface area and excluded volume in proteins: analytical equations for overlapping spheres and implications for the hydrophobic effect. J. Mol. Biol.* **178**, 63–89.

Richmond, T. J. & Richards, F. M. (1978). *Packing of alpha-helices: geometrical constraints and contact areas*. *J. Mol. Biol.* **119**, 537–555.

Rossmann, M. G. (1989). *The canyon hypothesis. J. Biol. Chem.* **264**, 14587–14590.

Rossmann, M. G. & Palmenberg, A. C. (1988). *Conservation of the putative receptor attachment site in picornaviruses. Virology*, **164**, 373–382.

Rowland, R. S. & Taylor, R. (1996). *Intermolecular nonbonded contact distances in organic crystal structures: comparison with distances expected from van der Waals radii*. *J. Phys. Chem.* **100**, 7384–7391.

Sgro, J.-Y. (1996). *Virus visualization*. In *Encyclopedia of virology plus* (CD-ROM version), edited by R. G. Webster & A. Granoff. San Diego: Academic Press.

Sharp, K. A., Nicholls, A., Fine, R. F. & Honig, B. (1991). *Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. Science*, **252**, 107–109.

Sherry, B., Mosser, A. G., Colonno, R. J. & Rueckert, R. R. (1986). *Use of monoclonal antibodies to identify four neutralization immunogens on a common cold picornavirus, human rhinovirus 14. J. Virol.* **57**, 246–257.

Sherry, B. & Rueckert, R. (1985). *Evidence for at least two dominant neutralization antigens on human rhinovirus 14. J. Virol.* **53**, 137–143.

Shrake, A. & Rupley, J. A. (1973). *Environment and exposure to solvent of protein atoms. Lysozyme and insulin*. *J. Mol. Biol.* **79**, 351–371.

Sibbald, P. R. & Argos, P. (1990). *Weighting aligned protein or nucleic acid sequences to correct for unequal representation*. *J. Mol. Biol.* **216**, 813–818.

Singh, R. K., Tropsha, A. & Vaisman, I. I. (1996). *Delaunay tessellation of proteins: four body nearest-neighbor propensities of amino acid residues*. *J. Comput. Biol.* **3**, 213–222.

Sreenivasan, U. & Axelsen, P. H. (1992). *Buried water in homologous serine proteases*. *Biochemistry*, **31**, 12785–12791.

Tanford, C. (1997). *How protein chemists learned about the hydrophobicity factor. Protein Sci.* **6**, 1358–1366.

Tanford, C. H. (1979). *Interfacial free energy and the hydrophobic effect*. *Proc. Natl Acad. Sci. USA*, **76**, 4175–4176.

Ten Eyck, L. F. (1977). *Efficient structure-factor calculation for large molecules by the fast Fourier transform. Acta Cryst.* A**33**, 486–492.

Tsai, J., Gerstein, M. & Levitt, M. (1996). *Keeping the shape but changing the charges: a simulation study of urea and its isosteric analogues*. *J. Chem. Phys.* **104**, 9417–9430.

Tsai, J., Gerstein, M. & Levitt, M. (1997). *Estimating the size of the minimal hydrophobic core*. *Protein Sci.* **6**, 2606–2616.

Tsai, J., Taylor, R., Chothia, C. & Gerstein, M. (1999). *The packing density in proteins: standard radii and volumes*. *J. Mol. Biol.* **290**, 253–266.

Tsai, J., Voss, N. & Gerstein, M. (2001). *Voronoi calculations of protein volumes: sensitivity analysis and parameter database*. *Bioinformatics*. In the press.

Voronoi, G. F. (1908). *Nouvelles applications des paramétres continus à la théorie des formes quadratiques*. *J. Reine Angew. Math.* **134**, 198–287.

Williams, M. A., Goodfellow, J. M. & Thornton, J. M. (1994). *Buried waters and internal cavities in monomeric proteins*. *Protein Sci.* **3**, 1224–1235.

Wodak, S. J. & Janin, J. (1980). *Analytical approximation to the accessible surface areas of proteins. Proc. Natl Acad. Sci. USA*, **77**, 1736–1740.

Xie, Q. & Chapman, M. S. (1996). *Canine parvovirus capsid structure, analyzed at 2.9 Å resolution. J. Mol. Biol.* **264**, 497–520.

Zhou, G., Somasundaram, T., Blanc, E., Parthasarathy, G., Ellington, W. R. & Chapman, M. S. (1998). *Transition state structure of arginine kinase: implications for catalysis of bimolecular reactions. Proc. Natl Acad. Sci. USA*, **95**, 8449–8454.

22.2

Adman, E., Watenpaugh, K. D. & Jensen, L. H. (1975). *N—H···S hydrogen bonds in Peptococcus aerogenes ferredoxin, Clostridium pasteurianum rubredoxin and Chromatium high potential iron protein*. *Proc. Natl Acad. Sci. USA*, **72**, 4854–4858.

Alber, T., Dao-pin, S., Wilson, K., Wozniak, J. A., Cook, S. P. & Matthews, B. W. (1987). *Contributions of hydrogen bonds of Thr 157 to the thermodynamic stability of phage T4 lysozyme. Nature (London)*, **330**, 41–46.

Artymiuk, P. J. & Blake, C. C. F. (1981). *Refinement of human lysozyme at 1.5 Å resolution. Analysis of non-bonded and hydrogen-bonded interactions. J. Mol. Biol.* **152**, 737–762.

Baker, E. N. (1995). *Solvent interactions with proteins as revealed by X-ray crystallographic studies.* In *Protein–solvent interactions*, edited by R. B. Gregory, pp. 143–189. New York: Marcel Dekker Inc.

Baker, E. N. & Hubbard, R. E. (1984). *Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol.* **44**, 97–179.

Blundell, T., Barlow, D., Borkakoti, N. & Thornton, J. (1983). *Solvent-induced distortions and the curvature of α-helices. Nature (London)*, **306**, 281–283.

Bordo, D. & Argos, P. (1994). *The role of side-chain hydrogen bonds in the formation and stabilization of secondary structure in soluble proteins. J. Mol. Biol.* **243**, 504–519.

Burley, S. K. & Petsko, G. A. (1986). *Amino–aromatic interactions in proteins. FEBS Lett.* **203**, 139–143.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996). *Crystal structure of a group I ribozyme domain: principles of RNA packing. Science*, **273**, 1678–1685.

Derewenda, Z. S., Derewenda, U. & Kobos, P. (1994). *(His)Cε—H···O=C< hydrogen bond in the active site of serine hydrolases. J. Mol. Biol.* **241**, 83–93.

Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). *The occurrence of C—H···O hydrogen bonds in proteins. J. Mol. Biol.* **252**, 248–262.

Edwards, R. A., Baker, H. M., Whittaker, M. M., Whittaker, J. W. & Baker, E. N. (1998). *Crystal structure of Escherichia coli manganese superoxide dismutase at 2.1 Å resolution. J. Biol. Inorg. Chem.* **3**, 161–171.

Fersht, A. R. & Serrano, L. (1993). *Principles in protein stability derived from protein engineering experiments. Curr. Opin. Struct. Biol.* **3**, 75–83.

Fersht, A. R., Shi, J.-P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Waye, M. M. Y. & Winter, G. (1985). *Hydrogen bonding and biological specificity analysed by protein engineering. Nature (London)*, **314**, 235–238.

Flocco, M. M. & Mowbray, S. L. (1995). *Strange bedfellows: interactions between acidic side-chains in proteins. J. Mol. Biol.* **254**, 96–105.

Gregoret, L. M., Rader, S. D., Fletterick, R. J. & Cohen, F. E. (1991). *Hydrogen bonds involving sulfur atoms in proteins. Proteins Struct. Funct. Genet.* **9**, 99–107.

Hagler, A. T., Huler, E. & Lifson, S. (1974). *Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. J. Am. Chem. Soc.* **96**, 5319–5327.

Harper, E. T. & Rose, G. D. (1993). *Helix stop signals in proteins and peptides: the capping box. Biochemistry*, **32**, 7605–7609.

Huggins, M. L. (1971). *50 years of hydrogen bonding theory. Angew. Chem. Int. Ed. Engl.* **10**, 147–208.