

may be understood in far greater detail with information from reflection analysis. A surprise has been just how ordered macromolecular crystals can be. This offers potential in new phasing methods such as multiple beam diffraction³³ and the exploitation of the coherent radiation opportunities available at third-generation synchrotron sources.

Acknowledgments

We thank Drs. Jeff Lovelace and Ardeschir Vahedi-Faridi (Toledo), Titus Boggon (New York), Peter Siddons (Brookhaven), and Prof. John Helliwell (Manchester) for technical and scientific contributions. This work was funded by NASA Grants NAG8-1380, NAG8-1825, and NAG-1836. E.H.S. is contracted to NASA through BAE-SYSTEMS Analytical Solutions.

³³E. Weckert and K. Hummer, *Acta Crystallogr. A* **53**, 108 (1997).

[16] Protein Structures at Atomic Resolution

By ZBIGNIEW DAUTER

Introduction

In recent years there has been a dramatic increase in the number of X-ray crystal structures of proteins refined at atomic resolution. This trend has been anticipated¹⁻³ and exceeds even the growth of the number of all protein structures deposited at the Protein Data Bank (PDB) (Fig. 1). There is no doubt that this explosion of atomic resolution structures is owed mainly to the advances in macromolecular crystallography methodology. The most important advances have been in the practice of crystal growth. The availability of convenient and quick protein-purification methods, efficient crystal-growth screening conditions, convenient crystallization chambers that employ only small amounts of sample, and sometimes mechanisms (robots) to automate the setting up of crystallization trials all contribute to these successes. A number of atomic resolution data have been obtained from crystals grown in microgravity, which makes it possible to use efficiently the full diffraction potential of very-high-quality protein crystals. The availability of bright synchrotron beam lines,

¹ Z. Dauter, V. S. Lamzin, and K. S. Wilson, *Curr. Opin. Struct. Biol.* **3**, 784 (1995).

² Z. Dauter, V. S. Lamzin, and K. S. Wilson, *Curr. Opin. Struct. Biol.* **7**, 681 (1997).

³ S. Longhi, M. Czjzek, and C. Cambillau, *Curr. Opin. Struct. Biol.* **8**, 730 (1998).

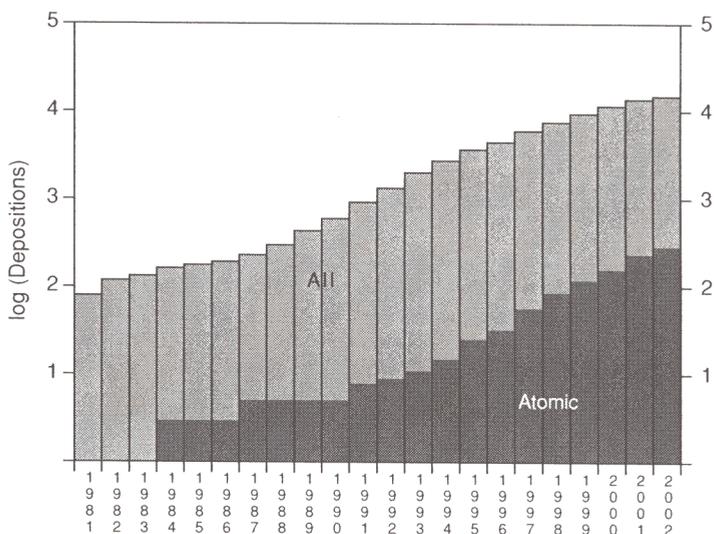


FIG. 1. A histogram showing (in logarithmic scale) the number of all atomic resolution protein structures available in the PDB at the end of each year.

maturation of cryopreservation techniques, and advances in the data processing and handling software, coupled with a general increase in computer speed and memory, is also responsible for this trend to a large extent.

Another factor is the increasing interest and appreciation of the researchers involved that proteins refined at atomic resolution give a wealth of additional, detailed information about the chemistry of the molecules being investigated. It is generally accepted that macromolecular crystallography provides valuable biochemical information about interactions between functional groups within a single protein or between partners in multimolecular complexes. It is not so generally appreciated that chemical reactions or interactions, which are crucial to all biological processes, depend on the precise alignment of particular atoms in their particular environment. The more accurately the stereochemical and electronic details about the interacting functional groups or molecules are known, the more intricate is the information about the chemistry that can be gained, be it an enzymatic reaction, or electron transfer, protein–ligand, protein–protein, or protein–nucleic acid recognition or signal transduction.

Definition of Atomic Resolution

There is no sharply defined criterion for the data resolution to become “atomic.” The primary result of the X-ray diffraction experiment is an

electron density map. The atomic model serves only as a representation of the electron density in terms more familiar to chemists. It is possible to build models consisting of individual atoms even at low resolution, since the building “bricks” of macromolecules, i.e., amino acids and nucleotides, are well known. The knowledge of bond lengths and angles within those bricks is used in the interpretation of the electron density maps, and it is applied in the form of constraints or restraints during refinement of the atomic model. At low resolution the individual atoms are not represented by well-resolved peaks in the Fourier syntheses, but when the number of the measured reflection intensities (observables) increases, i.e., the resolution of the diffraction data improves, the electron density tends to show resolved peaks corresponding to individual atoms. This process is gradual, and even at the resolution poorer than 2 Å individual peaks can appear in the well-defined parts of the model. On the other hand, even at a resolution better than 1 Å, some atoms in the flexible parts of a macromolecule may be unresolved in the map.

The current consensus among crystallographers is that diffraction data are considered “atomic” if the resolution reaches at least down to d -spacings of 1.2 Å, and at this limit a majority of intensities are above the 2σ level. This condition has been proposed by Sheldrick⁴ on the basis of his experience in solving large structures by direct methods, which are based on the atomicity principle that requires that peaks in the Fourier synthesis corresponding to individual atoms do not overlap. The 1.2-Å resolution limit is a little shorter than the average bond length within crystallized molecules, so that the requirement of nonoverlapping atoms is fulfilled.

In small molecule crystallography the diffraction data from organic crystals usually have been collected on a four-circle diffractometer to the maximum limit of the 2θ angle achievable with the copper anode source, which corresponds to about 0.8 Å resolution. For noncentrosymmetric structures this gives about five measured intensities per one parameter of the anisotropically refined model. Since proteins contain a large amount of bulk solvent, the same ratio of observables per parameter is achieved at a resolution of about 1.0 Å, which confirms the validity of a similar limit for the macromolecular diffraction data to be classified as atomic.

Atomic Resolution Data Collection

Collection of data from crystals diffracting to atomic resolution basically does not differ significantly from the “ordinary” procedure.⁵⁻⁷ However, attention should be paid to certain points. Atomic resolution

⁴ G. M. Sheldrick, *Acta Crystallogr. A* **46**, 467 (1990).

data should be complete at both highest and lowest resolution limits. The highest resolution reflections carry the information about the finest details of the structure, whereas the strong, low-resolution reflections modulate the Fourier maps most significantly, and lack of them will seriously impair the interpretation of electron density maps. Reflection intensities at the high-resolution limit of diffraction are very weak in comparison to the strong, low-resolution reflections. To obtain a meaningful measurement of those weakest intensities, exposure time has to be long. Since all two-dimensional detectors, in particular charge-coupled device (CCD)-based detectors and imaging plates, have limited dynamic range, many strongest reflection profiles will have overloaded pixels and therefore will not be properly measured with such long exposure times. To estimate the strongest intensities properly, one must make a separate pass of rotation images with much shorter exposures (or attenuated X-ray beam). Sometimes more than two passes may be advised, since the effective exposure difference between a pair of consecutive passes should not exceed about 10 for successful scaling of all data. The short-exposure data collection pass should encompass all overloaded reflections from the previous one, and its resolution limit should ensure enough reflections common to both passes for the scaling procedure to be effective. Therefore it can be performed with a longer crystal-to-detector distance and wider oscillation ranges.

If a crystal has relatively long cell dimensions, reflection profiles at high diffraction angles at the edge of the detector may tend to overlap, especially when the longest cell dimension is oriented parallel to the X-ray beam. If the crystal mosaicity is substantial, reflections will overlap even with a very thin oscillation range. It is most beneficial to orient the crystal with its longest (primitive) cell dimension along the detector spindle axis, since then it will never come to be parallel to the beam. Often this can be achieved at κ -geometry goniostats.

As for any data collection by the rotation method, proper limits of total rotation range should be selected, to ensure that at least the complete asymmetric unit of the reciprocal space is covered. However, it is beneficial to collect more data than the minimum, since increased redundancy leads to more accurate estimation of the symmetry-averaged intensities.

In the scaling and merging procedure, attention should be directed to the proper estimation of intensity uncertainties (sigmas). They are used in the atomic model refinement as reflection weights, and if badly

⁵ Z. Dauter, *Methods Enzymol.* **276**, 326 (1997).

⁶ Z. Dauter, *Acta Crystallogr. D* **55**, 1703 (1999).

⁷ Z. Dauter and K. S. Wilson, in "International Tables of Crystallography," Vol. F, p. 177. Kluwer Academic Publishers, Dordrecht, 2001.

estimated they will adversely influence the refinement procedure. Bad error estimates will lead to biased values of refined parameters and their standard uncertainties, and therefore to a wrong goodness of fit. Since two-dimensional (2-D) detectors do not measure individual X-ray quanta, sigmas estimated according to the counting statistics need to be corrected. All data collection programs have means to do that, based on the t -plot or the χ -square test.

Refinement Software and Protocols

Several programs can be used to refine protein models against atomic resolution data. By far the most popular has been SHELXL.⁸ It was created by George Sheldrick, based on an enormous amount of experience with refining small structures. The program has been extended to include many options intended especially for macromolecules. It is extremely flexible, allowing the user to design various protocols for anisotropic refinement, restraints, rigid body groups, H-atom positioning, automatic solvent selection, etc., but also providing well-validated default values for all refinement parameters. Another program, gaining increasing popularity and providing similar refinement options, is REFMAC.⁹

When atomic-resolution data are available, it may be advisable to use only limited resolution (1.5–1.8 Å) in the first stages of refinement of the isotropic model with the conjugate-gradient option. It will speed up the procedure considerably, since at twice higher resolution the number of reflections increases 8-fold. These moderate-resolution reflections within an atomic-resolution data set should be strong and well measured, and will produce relatively high-quality electron density maps, so that many features such as more pronounced double conformations of side chains, well-defined solvent atoms, etc. can be modeled confidently.

After extending the resolution to the full limit, the model can be refined anisotropically and hydrogen atoms can be included at their calculated positions. More subtle features appearing in the electron density can then be modeled and the refinement process continued to convergence.

Geometric and B -factor restraints are necessary only to preserve acceptable stereochemistry of more flexible parts of the model. At atomic resolution there is a large excess of observables over the refined parameters, and the influence of restraints on well-defined atoms is minimal.

⁸ G. M. Sheldrick and T. R. Schneider, *Methods Enzymol.* **277**, 319 (1997).

⁹ G. N. Murshudov, A. A. Vagin, A. Lebedev, K. S. Wilson, and E. J. Dodson, *Acta Crystallogr. D* **55**, 792 (1999).

However, releasing them completely will lead to large movements of atoms in poorly defined parts of the model.

One of the most difficult features, even at atomic resolution, remains the poorly defined solvent molecules. Water molecules in the first solvent shell usually behave well, but further away from the protein surface or near its flexible regions, electron density peaks may be confusing. Sometimes it is possible to assign partially occupied water molecules in such peaks, or even to refine their occupancies within alternative hydrogen bonding networks or coupled to the protein side chains in double conformation. However, even when one works at atomic resolution, the difficult problem of modeling weak solvent sites still remains, although it is shifted further away from the protein surface.

It is a customary procedure to set aside a fraction (usually 5%) of all reflections throughout the entire refinement process for cross-validation by R_{free} .¹⁰ This is perhaps less necessary with atomic data than at moderate or low resolution, where the danger of overfitting the model is more severe. The R_{free} is extremely useful for validating the general refinement protocols but, as a global factor, cannot serve to validate particular subtle features such as individual side chain double conformations or weak water sites.¹¹ The refinement protocols at atomic resolution are rather well established,¹² and there is no need to validate anisotropic atomic displacements or inclusion of hydrogen atoms. The individual structural features have to be validated by the (preferably unbiased) electron density maps and by chemical sense. Even if R_{free} is used to monitor the progress of refinement, it is enough to assign about 1000 reflections for this purpose, which is usually significantly less than 5% of the total data, and use the rest for refinement. In the last cycle of the refinement procedure all reflections should be used to obtain the final statistics. In this sense there should be no "final R_{free} value."

Accuracy of Coordinates and Stereochemistry

The precision of the refined atomic positional and displacement parameters depends mainly on the number of reflection intensities used, taking into account the accuracy of their estimation, and on the agreement between observed structure amplitudes and those calculated from the current atomic model. At data resolution lower than atomic, it is possible to

¹⁰ A. T. Brünger, *Nature* **355**, 472 (1992).

¹¹ E. J. Dodson, G. J. Kleywegt, and K. S. Wilson, *Acta Crystallogr. D* **52**, 228 (1996).

¹² Z. Dauter, G. N. Murshudov, and K. S. Wilson, in "International Tables of Crystallography," Vol. F, p. 393. Kluwer Academic Publishers, Dordrecht, 2001.

estimate only roughly the average error of atomic positions. It is obvious that within the protein model some parts are better defined than other, more flexible regions, and such differences can be inferred only indirectly from the atomic B -factors. There are global precision indicators that are traditionally used at moderate resolution; these are derived from the Luzzati plot,¹³ σ_A ,¹⁴ or, more recently, the diffraction precision indicator (DPI).^{15,16} These are not completely appropriate at atomic resolution, since the least-squares refinement procedure offers the proper way of estimating errors of the individual parameters from the inversion of the least-squares matrix. From these, error estimates of all derived quantities, such as bond lengths and angles, can be obtained. In practice it is possible to perform the inversion of the full least-squares matrix only for smaller structures, having 100–200 atoms. For larger models it is advisable to use the block-matrix least-squares approximation in the last cycles, including in each cycle a region containing 15–20 amino acids (about 1500 parameters), and simultaneously removing all restraints and damping all parameter shifts to zero.⁸ This is a satisfactory compromise, providing individual error estimates, perhaps with slightly less confidence. However, for this procedure to be valid, uncertainties of the measured diffraction intensities have to be estimated properly.

The precision of atomic coordinates in well-defined parts of protein models refined at atomic resolution is about 0.02 Å or even better. Coordinates of atoms heavier than carbon, nitrogen, and oxygen, such as sulfur or metals, may be smaller than 0.01 Å. Such accuracy can be reached when an R -factor falls to 10–12%. Of course, flexible parts of the protein model, with high B -factors, will be much less accurately defined.

Atomic Resolution Structures in the PDB

Table I lists all protein structures deposited in the Protein Data Bank¹⁷ before end of the year 2002 that were refined at a resolution at least 1.2 Å. Since there is no sharp definition when a polypeptide can be accepted as a protein, all structures built from amino acids are listed, including relatively small oligopeptide antibiotics. The table gives the PDB code, protein size, resolution, R -factor, deposition date, and the appropriate reference.

¹³ P. V. Luzzati, *Acta Crystallogr.* **5**, 802 (1952).

¹⁴ R. Read, *Acta Crystallogr. A* **42**, 140 (1986).

¹⁵ D. W. J. Cruickshank, *Acta Crystallogr. D* **55**, 583 (1999).

¹⁶ D. M. Blow, *Acta Crystallogr. D* **58**, 792 (2002).

¹⁷ H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *Nucleic Acids Res.* **28**, 235 (2000).

The structures are sorted according to the deposition date. It is apparent that until 1997 there were only a few atomic resolution models deposited per year, but after 1998 the number of atomic resolution structures grew very fast. This parallels the recent progress in the X-ray methodology, such as the availability of bright synchrotron beam lines, fast, sensitive, and accurate detectors, efficient software for data processing, model refinement, and graphics inspection.

In spite of the availability of several user-friendly and sophisticated programs, the task of refining and properly validating atomic resolution models is complicated and time consuming. Very fine features can be confidently validated only by inspection of the atomic model and electron density maps on the graphics display and manual intervention. Therefore the time elapsed from atomic data collection to structure deposition and publication is usually several months, if not years. The contents of Table I are therefore unavoidably outdated, and taking into account the current trend, it may be expected that in the future the number of available atomic resolution structures will grow even faster. This will provide a firmer basis for more statistically valid discussion of various subtle features of protein molecules as seen in the crystalline state. In the subsequent sections some of these fine features will be presented.

Atomic Resolution Features

Various fine features become apparent at atomic resolution. Some of them may be visible even at lower resolution, but when the resolution approaches 1 Å, they may be modeled with higher confidence.

Multiple Conformations

Double conformations of amino acid side chains can be modeled at moderate resolution, even lower than 2 Å, but at atomic resolution they become more apparent. It is not uncommon to see up to 10–15% of all side chains displaying multiple rotamers. This phenomenon is not restricted to residues at the protein surface; often buried amino acids form clusters with concerted alternative sets of side chain rotamers. Sometimes residues in the free enzyme active sites are also observed in multiple conformations.¹⁸ At the protein surface the double conformations of side chains are usually coupled with partially occupied water sites. The split atomic sites can be refined with the common occupancy x for one conformation and $1-x$ for the alternative one.

¹⁸ D. Ghosh, M. Sawicki, P. Lala, M. Erman, W. Pangborn, J. Eyzaguirre, R. Gutierrez, H. Jornvall, and D. J. Thiel, *J. Biol. Chem.* **276**, 11159 (2001).

TABLE I
 ATOMIC RESOLUTION STRUCTURES OF PROTEINS IN THE PDB AVAILABLE AT THE END OF 2002, SORTED ACCORDING TO THE DEPOSITION DATE

PDB code	SF availability	Protein	Residues	Atoms	Resolution (Å)	R-factor	Deposition date	Release date	References ^a
5PTI	-	BPTI	58	752	1.0	0.200(I)	5-10-1984	29-10-1984	1
4RXN	-	Rubredoxin	54	550	1.2	0.128	15-10-1984	1-04-1985	2
5RXN	+	Rubredoxin combined refinement	54	525	1.2	0.115	15-10-1984	1-04-1985	3
1XY1	-	Oxytocin wet form	20	149	1.04	0.088	5-06-1987	16-04-1988	4
1XY2	-	Oxytocin dry form	10	75	1.20	0.100	5-06-1987	16-04-1988	4
8RXN	+	Rubredoxin	52	497	1.0	0.147(I)	26-08-1991	31-10-1993	5
1CBN	-	Crambin	46	344	0.83	0.106	11-10-1991	31-01-1994	6
1IFC	-	Fatty acid-binding protein	132	1295	1.19	0.169(I)	19-12-1991	31-01-1994	7
2SN3	-	Scorpion toxin	65	625	1.2	0.192(I)	20-02-1992	31-01-1994	8
1ARB	-	Protease I	268	2077	1.2	0.149	15-04-1993	31-10-1993	9
1CNR	-	Crambin	46	330	1.05	0.095	15-07-1993	31-04-1994	10
1ETL	-	Enterotoxin	13	94	0.89	0.073	15-03-1994	29-01-1996	11
1ETM	-	Enterotoxin	13	95	0.89	0.065	15-03-1994	29-01-1996	11
1ETN	-	Enterotoxin	13	95	0.89	0.088	15-03-1994	29-01-1996	12
1IGD	+	Protein G	61	588	1.1	0.193	5-08-1994	1-11-1994	13
1BPI	+	BPTI	58	626	1.1	0.146	18-02-1995	3-06-1995	14
1RGE	+	Ribonuclease Sa 2GMP	192	1968	1.15	0.109	5-06-1995	14-10-1996	15
1RGF	+	Ribonuclease Sa iso	192	1887	1.2	0.167	5-06-1995	14-10-1996	15
1RGG	+	PROLSQ Ribonuclease Sa aniso	192	1840	1.2	0.106	5-06-1995	14-10-1996	15
1RGH	+	SHELXL Ribonuclease Sa iso	192	1794	1.2	0.173	5-06-1995	14-10-1996	15
1CTJ	+	SHELXL Cytochrome c ₆	89	853	1.10	0.140	8-08-1995	10-06-1996	16
1LKK	-	SH2 domain	110	1051	1.0	0.133	10-11-1995	8-03-1996	17

IIRN	+	Rubredoxin Zn	54	501	1.2	0.103	13-12-1995	3-04-1996	18
IIRO	+	Rubredoxin Fe	54	526	1.1	0.090	13-12-1995	3-04-1996	18
2ERL	+	Pheromone Er-1	40	327	1.0	0.129	20-12-1995	11-07-1996	19
IPEN	-	α -Conotoxin	17	122	1.1	0.127	29-01-1996	21-04-1997	20
1AMM	-	γ B-Crystallin	174	1866	1.2	0.184(I)	20-03-1996	8-11-1996	21
1NOT	-	α -Conotoxin	14	118	1.2	0.178	2-05-1996	7-12-1996	22
1JET	-	OppA	520	4765	1.2	0.229	3-07-1996	15-05-1997	23
1IXG	-	Phosphate-binding protein mutant	321	2444	1.05	0.112	1-08-1996	4-02-1998	24
1IXH	-	Phosphate-binding protein	321	2443	0.98	0.117	1-08-1996	4-02-1998	24
1JBC	+	Concanavalin	237	2098	1.2	0.142	23-08-1996	12-02-1997	25
2KNT	+	Kunitz domain	58	515	1.20	0.149	15-01-1997	15-07-1997	26
1AA5	+	Vancomycin	14	258	0.90	0.112	23-01-1997	20-08-1997	27
1NLS	+	Concanavalin A	237	2130	0.94	0.127	28-01-1997	26-11-1997	28
1AB1	+	Crambin	46	329	0.89	0.147	31-01-1997	12-08-1997	29
1CEX	+	Cutinase	214	1704	1.0	0.094	18-02-1997	20-08-1997	30
2PTH	-	Peptidyl tRNA hydrolase	193	1661	1.2	0.196(I)	25-03-1997	25-03-1998	31
3LZT	+	Lysozyme triclinic low temperature	129	1275	0.92	0.092	23-03-1997	25-03-1998	32
4LZT	+	Lysozyme triclinic room temperature	129	1163	0.95	0.108	31-03-1997	1-04-1998	32
1AGY	+	Cutinase	200	1710	1.15	0.175(I)	26-03-1997	1-04-1998	30
1AHO	+	Scorpion toxin	64	629	0.96	0.158	8-04-1997	15-10-1997	33
2IGD	+	Protein G	61	574	1.1	0.094	30-04-1997	29-06-1998	34
1AKG	+	α -Conotoxin Pnib	17	134	1.1	0.147(I)	18-05-1997	20-05-1998	35
1ALX	+	Gramicidin D methanol	35	324	1.20	0.089	5-06-1997	4-03-1998	36
1ALZ	+	Gramicidin D ethanol	36	332	0.86	0.164	6-06-1997	4-03-1998	36
1AL4	+	Gramicidin D propanol	36	328	1.13	0.067	11-06-1997	4-03-1998	36
1SHO	+	Vancomycin	12	316	1.09	0.105	23-07-1997	24-12-1997	37
1ATG	+	Molybdate-binding protein	231	2098	1.2	0.164(I)	14-08-1997	14-10-1998	38
1MRO	-	Methyl CoM reductase	2474	21038	1.16	0.197(I)	1-10-1997	11-11-1998	39
2FDN	+	Ferredoxin	55	490	0.94	0.100	1-10-1997	8-04-1998	40

(continued)

TABLE I (continued)

PDB code	SF availability	Protein	Residues	Atoms	Resolution (Å)	R-factor	Deposition date	Release date	References ^d
ILKS	+	Lysozyme hen trichin	129	1222	1.1	0.110	9-10-1997	29-04-1998	41
INKD	+	ROP mutant	65	582	1.09	0.101	23-09-1997	23-03-1999	42
IPSR	+	Psoriasis	200	1975	1.05	0.109	27-11-1997	13-01-1999	43
1A0M	+	α -Conotoxin	34	284	1.1	0.134	3-12-1997	13-01-1999	44
1RB9	+	Rubredoxin	53	490	0.92	0.073	21-12-1997	16-02-1999	34
1JSE	+	Lysozyme turkey	129	1141	1.12	0.104	5-01-1998	29-04-1998	45
1JSF	+	Lysozyme human	130	1207	1.15	0.115	5-01-1998	29-04-1998	45
1A6G	+	Myoglobin CO	151	1440	1.15	0.124	25-02-1998	21-10-1998	46
1A6K	+	Myoglobin MetH ₂ O	151	1446	1.1	0.128	26-02-1998	6-04-1999	46
1A6M	+	Myoglobin O ₂	151	1445	1.0	0.119	26-02-1998	6-04-1999	46
1A6N	+	Myoglobin deoxy	151	1443	1.15	0.119	26-02-1998	6-04-1999	46
1A7S	+	Heparin-binding protein	221	2083	1.12	0.159	17-03-1998	23-03-1999	47
1A7Y	+	Actinomycin D	33	314	0.94	0.056	19-03-1998	23-03-1999	48
1A7Z	+	Actinomycin Z3	22	307	0.95	0.079	19-03-1998	23-03-1999	48
3SIL	-	Sialidase	379	3527	1.05	0.116	7-07-1998	11-11-1998	49
1BKR	+	Calponin homology domain	109	1095	1.1	0.141	10-07-1998	15-07-1998	50
3PYP	+	Photoactive yellow protein	125	1124	0.85	0.125	28-07-1998	1-06-1999	51
7A3H	-	Endoglucanase Cel5A	303	2869	0.95	0.110	5-08-1998	6-08-1999	52
1BQ8	+	Rubredoxin Met mutant	54	612	1.10	0.119	22-08-1988	26-08-1998	53
1BQ9	+	Rubredoxin FmMet mutant	54	583	1.20	0.137	22-08-1998	26-08-1998	53
1BRF	+	Rubredoxin	53	600	0.95	0.132	24-08-1998	2-09-1998	53
1MUN	-	MutY mutant	225	2140	1.20	0.124	26-08-1998	26-08-1999	54
1BS9	-	Acetylxylin esterase	207	1590	1.1	0.124	1-09-1998	18-05-1999	55
1GCI	-	Subtilisin <i>B. lentus</i>	269	2277	0.78	0.099	2-09-1998	21-10-1998	56
1TAX	+	Xylanase I	302	2470	1.14	0.180(I)	15-09-1998	16-09-1999	57
2PVB	+	Parvalbumin	107	1025	0.91	0.110	2-10-1998	7-10-1998	58
1BXO	+	PP4-penicillopepsin	323	2981	0.95	0.099	7-10-1998	14-10-1998	59

1SWU	+	Streptavidin mutant	508	4085	1.14	0.126	12-10-1998	10-11-1999	60
1BX7	+	Hirustasin	55	428	1.2	0.179	14-10-1998	27-04-1999	61
1BY1	+	Dethiobiotin	224	2093	0.97	0.116	15-10-1998	15-06-1999	62
1BYZ	+	Alpha-1 peptide	52	479	0.90	0.085	20-10-1998	28-10-1998	63
3AL1	+	α -Helical peptide	26	238	0.75	0.131	26-10-1998	4-11-1998	64
1BZP	+	Myoglobin deoxy	153	1501	1.15	0.114	2-11-1998	10-05-1999	65
1BZR	+	Myoglobin CO	153	1478	1.15	0.124	3-11-1998	10-05-1999	65
1BZ6	+	Myoglobin MetH ₂ O	153	1503	1.2	0.090	5-11-1998	11-11-1998	65
2NLR	+	Endoglucanase CelB2	234	1966	1.2	0.112	2-11-1998	10-11-1999	66
1B0Y	+	HiPIP mutant	85	837	0.93	0.155	15-11-1998	16-12-1998	67
1B6G	-	Haloalkane dehalogenase	310	3122	1.15	0.105	14-01-1999	13-07-1999	68
8A3H	-	Endoglucanase Cel5A complex	303	2952	0.97	0.104	20-01-1999	11-01-2000	69
1B9O	+	α -Lactalbumin	123	1134	1.15	0.119	14-02-1999	31-03-1999	70
1CC7	-	AtxI metallochaperone apo	73	691	1.20	0.145	4-03-1999	12-12-1999	71
1CC8	-	AtxI metallochaperone Hg	73	703	1.02	0.146	4-03-1999	12-12-1999	71
1MFM	+	SOD mutant	153	1394	1.02	0.118	16-04-1999	21-04-1999	72
1CKU	+	HiPIP	170	1598	1.2	0.128	24-04-1999	13-05-1999	67
1V FY	+	FYVE domain	73	643	1.15	0.174(I)	26-04-1999	6-05-1999	73
1Q1O	+	HEW lysozyme	129	1182	1.20	0.147	14-06-1999	11-04-2001	74
1QQ4	-	α -Lytic protease mutant	198	1656	1.20	0.174(I)	10-06-1999	15-06-1999	75
1QRW	-	α -Lytic protease mutant	198	1649	1.20	0.177(I)	16-06-1999	18-06-1999	75
1QJ4	+	Hydroxynitrile lyase	257	2627	1.1	0.116	21-06-1999	10-10-1999	76
1QTN	-	Caspase 8	264	2319	1.2	0.166	28-06-1999	20-09-2000	77
1QTW	-	Endonuclease IV	285	2586	1.02	0.124	29-06-1999	31-08-1999	78
1Q09	+	yigF	384	3315	1.2	0.164	7-07-1999	1-12-1999	79
1C0P	+	D-Amino acid oxidase	363	3466	1.20	0.118	19-07-1999	22-11-2000	80
1C9O	+	Cold shock protein	132	1338	1.17	0.125	3-08-1999	2-04-2000	81
1OL0	+	Endonuclease	482	4541	1.10	0.128	18-08-1999	7-05-2000	82
1CXO	+	ASV integrase ammonium sulfate	162	1320	1.02	0.129	30-08-1999	8-09-1999	83

(continued)

TABLE I (continued)

PDB code	SF availability	Protein	Residues	Atoms	Resolution (Å)	R-factor	Deposition date	Release date	References ^d
1CZ9	+	ASV integrase citrate	162	1301	1.20	0.147	1-09-1999	8-09-1999	83
1CZB	+	ASV integrase HEPES	162	1238	1.06	0.143	1-09-1999	8-09-1999	83
1CZP	+	Ferredoxin <i>Anabaena</i>	196	1877	1.17	0.140	6-09-1999	14-01-2000	84
1QLW	-	Esterase	656	5588	1.1	0.143	17-09-1999	10-02-2000	85
1D2U	+	Nitrophorin 4	184	1721	1.15	0.130	28-09-1999	3-10-2001	86
1D4T	+	SAP protein	115	1151	1.10	0.127	6-10-1999	14-10-1999	87
1D5T	+	α -GDI	433	3815	1.04	0.172	11-10-1999	25-10-2000	88
1QNJ	+	Elastase	240	2177	1.1	0.127	15-10-1999	31-03-2000	89
1QOW	+	Mersacidin	120	825	1.06	0.141	9-11-1999	26-05-2000	90
1C5E	+	λ -Phage gpD	285	2394	1.1	0.098	18-11-1999	8-03-2000	91
1DJT	-	Scorpion toxin Bmk	128	1198	1.2	0.109	5-12-1999	6-12-2000	92
1DS1	-	Clavaminc acid synthase	324	2925	1.08	0.135	6-01-2000	6-07-2000	93
1DY5	+	Ribonuclease A	248	2340	0.87	0.101	26-01-2000	28-03-2000	94
1C75	+	Cytochrome c_{553}	71	663	0.97	0.120	9-02-2000	22-03-2000	95
1C7K	+	Zinc protease	132	1133	1.00	0.148	19-02-2000	25-04-2001	96
1EJG	+	Crambin	46	327	0.54	0.090	2-03-2000	5-04-2000	97
1E0W	-	Xylanase 10A	302	2385	1.2	0.09	10-04-2000	5-04-2000	98
1ET1	-	Human parathyroid hormone	68	684	0.90	0.137	12-04-2000	6-09-2000	99
1EUW	+	dUTPase	152	1315	1.05	0.141	17-04-2000	3-05-2000	100
1EXR	+	Calmodulin	148	1328	1.0	0.137	3-05-2000	20-09-2000	101
1F86	+	Transferrin mutant	230	2069	1.10	0.147	29-06-2000	29-06-2001	102
1F94	+	Bucandin	63	616	0.97	0.124	6-07-2000	26-07-2000	103
1F98	+	Yellow protein mutant T50V	125	1096	1.15	0.132	7-07-2000	21-07-2000	104
1F9I	+	Yellow protein mutant Y42F	125	1090	1.10	0.127	10-07-2000	21-07-2000	104
1E4M	-	Myrosinase	501	5131	1.20	0.124	10-07-2000	25-05-2001	105

1HET	+	Alcohol dehydrogenase,	748	7120	1.15	0.123	14-07-2000	31-05-2001	106
1FSG	+	HGPRT	466	4769	1.05	0.122	8-09-2000	6-12-2000	107
1FN8	+	Trypsin <i>Fusarium</i> ARG	227	2007	0.81	0.108	21-08-2000	7-02-2001	108
1FY4	+	Trypsin <i>Fusarium</i> GLN	227	1991	0.81	0.108	28-09-2000	7-02-2001	108
1FY5	+	Trypsin <i>Fusarium</i> LYS2	227	1967	0.81	0.124	28-09-2000	7-02-2001	108
1GDN	+	Trypsin <i>Fusarium</i> LYS	227	2036	0.81	0.108	28-09-2000	7-02-2001	108
1GDO	+	Trypsin <i>Fusarium</i> F	227	2107	0.93	0.099	28-09-2000	7-02-2001	108
1GDU	+	Trypsin <i>Fusarium</i> N	227	1942	1.07	0.104	29-09-2000	7-02-2001	108
1FXM	+	Xylanase I	302	2478	1.14	0.117	26-09-2000	7-03-2001	109
1FY2	+	Aspartyl dipeptidase iso	229	1877	1.20	0.218(1)	28-09-2000	10-01-2001	110
1FYE	-	Aspartyl dipeptidase aniso	229	1877	1.20	0.157	29-09-2000	10-01-2001	110
1E9G	+	Inorganic pyrophosphatase	572	5561	1.15	0.136	12-10-2000	19-03-2001	111
1G2B	+	SH3 domain	62	674	1.12	0.148	18-10-2000	1-11-2000	112
1G2Y	+	HNF-1 α mutant L12SeM	128	1039	1.00	0.195(1)	23-10-2000	17-01-2001	113
1G2Z	+	HNF-1 α mutant L13SeM	64	561	1.15	0.216(1)	23-10-2000	17-01-2001	113
1E9W	+	Thiostrepton	7	118	1.02	0.112	27-10-2000	23-03-2001	114
1G4I	+	bpPLA ₂	123	1246	0.97	0.094	27-10-2000	4-04-2001	115
1G66	-	Acetylxyylan esterase	207	1780	0.90	0.107	3-11-2000	17-01-2001	116
1G6X	+	BPTI mutant	58	663	0.86	0.107	8-11-2000	9-05-2001	117
1G7A	+	Human insulin	204	1875	1.20	0.169	9-11-2000	3-08-2001	118
1HDO	-	Biliverdin IX β reductase	206	1951	1.15	0.124	16-11-2000	28-02-2001	119
1HE2	-	Biliverdin IX β reductase	206	1922	1.20	0.122	18-11-2000	28-02-2001	119
1G8T	+	Endonuclease	490	4570	1.10	0.129	21-11-2000	6-12-2000	82
1HEU	+	Alcohol dehydrogenase, Cd	748	6983	1.15	0.119	25-11-2000	31-05-2001	106
1HF6	-	Endoglucanase Cel5A complex	303	2863	1.15	0.115	28-11-2000	29-11-2001	120
1GA6	+	PSCP	375	3219	1.00	0.099	29-11-2000	13-12-2000	121
1HG7	+	Antifreeze protein	66	565	1.15	0.120	13-12-2000	31-01-2001	122
1HQJ	-	Coiled-coil peptide	192	1810	1.20	0.170	18-12-2000	14-03-2001	123
1HVB	-	Dd peptidase, cell wall	349	3206	1.17	0.112	8-01-2001	7-02-2001	124
1HJ8	+	Trypsin, salmon	222	1916	1.00	0.118	9-01-2001	4-01-2002	125
1HJ9	+	Trypsin, bovine	229	1913	0.95	0.117	10-01-2001	4-01-2002	125

(continued)

TABLE I (continued)

PDB code	SF availability	Protein	Residues	Atoms	Resolution (Å)	R-factor	Deposition date	Release date	References ^a
1I27	-	Transcription factor IIF	73	737	1.02	0.126	7-02-2001	7-03-2001	126
1I2T	+	Hyperplastic discs protein	61	602	1.04	0.143	12-02-2001	18-04-2001	127
1I3H	+	Concanavalin A dimannose	237	2087	1.20	0.170	15-02-2001	25-07-2001	128
1I40	+	Inorganic pyrophosphatase CaPP	175	1636	1.10	0.117	19-02-2001	5-12-2001	129
1I6T	+	Inorganic pyrophosphatase Ca	175	1628	1.20	0.129	5-03-2001	5-12-2001	129
1I4U	+	α -Crustacyanin	362	3381	1.15	0.150	23-02-2001	19-09-2001	130
1H97	+	Trematode hemoglobin	294	2778	1.17	0.121	2-03-2001	21-06-2001	131
1I76	+	Neutrophil collagenase	163	1586	1.20	0.129	8-03-2001	21-03-2001	132
1I80	+	Cytochrome <i>c</i>	114	1148	1.15	0.151	15-03-2001	4-04-2001	133
1IC6	-	Proteinase K	279	2558	0.98	0.114	30-03-2001	11-04-2001	134
1IEE	+	HEW lysozyme	129	1219	0.94	0.123	9-04-2001	8-08-2001	135
1HBN	-	Coenzyme M reductase	2478	21620	1.16	0.124	19-04-2001	16-08-2001	136
1IJV	+	β -Defensin-1	72	735	1.20	0.161	30-04-2001	24-10-2001	137
1KOI	+	Nitrophorin 4	184	1737	1.08	0.117	3-05-2001	9-01-2002	86
1H4G	+	Xylanase from <i>B. agarradhaerens</i>	414	3920	1.10	0.158	11-05-2001	9-05-2002	138
1H4X	+	SpoIIAA	234	2212	1.16	0.133	15-05-2001	6-07-2001	139
1J98	-	Quorum sensing LuxS	157	1443	1.20	0.127	24-05-2001	6-06-2001	140
1H5V	+	Endoglucanase Cel5A complex	304	2955	1.10	0.121	28-05-2001	23-05-2002	141
1JBE	+	ApoCheY	128	1203	1.08	0.121	4-06-2001	8-08-2001	142
1JCJ	+	DERA mutant	520	4371	1.10	0.142	9-06-2001	31-10-2001	143
1JCL	+	DERA	520	4522	1.05	0.143	9-06-2001	31-10-2001	143
1EA7	+	Sphericase	310	2015	0.93	0.097	10-07-2001	4-07-2002	144
1JM1	+	Rieske protein soxF	204	1806	1.11	0.106	17-07-2001	17-07-2002	145
1JF8	+	Arsenate reductase	131	1263	1.12	0.208	20-06-2001	20-06-2002	146

1JFB	-	Nitric oxide reductase, ferric	404	4154	1.00	0.102	20-06-2001	20-12-2001	147
1JFC	-	Nitric oxide reductase, ferrous CO	404	3974	1.05	0.117	20-06-2001	20-12-2001	147
1JG1	+	Methyltransferase	235	1957	1.20	0.147	22-06-2001	16-11-2001	148
1JK3	-	Macrophage elastase	158	1478	1.09	0.169	11-07-2001	28-09-2001	149
1EB6	+	Deuterolysin	177	1608	1.00	0.104	19-07-2001	23-11-2001	150
1GKM	+	Histidine ammonia lyase	509	4326	1.00	0.119	16-08-2001	5-04-2002	151
1IQZ	+	Ferredoxin	81	791	0.92	0.097	30-08-2001	13-02-2002	152
1IR0	+	Ferredoxin	81	764	1.00	0.096	30-08-2001	13-02-2002	152
1JXT	+	Crambin at 160 K	46	330	0.89	0.145	8-09-2001	31-10-2001	153
1JXU	+	Crambin at 240 K	46	330	0.99	0.090	9-09-2001	31-10-2001	153
1JXW	+	Crambin at 180 K	46	330	0.89	0.137	10-09-2001	31-10-2001	153
1JXX	+	Crambin at 200 K	46	330	0.89	0.136	10-09-2001	31-10-2001	153
1JXY	+	Crambin at 220 K	46	330	0.89	0.145	10-09-2001	31-10-2001	153
1GMX	+	GlpE sulfurtransferase	108	982	1.10	0.128	25-09-2001	28-11-2001	154
1K1T	-	HIV-1 protease mutant	205	1744	1.20	0.196(I)	25-09-2001	10-07-2002	155
1K2A	+	Neurotoxin EDN	136	1352	1.00	0.132	26-09-2001	3-04-2002	156
1K4I	-	Dihydrobutanone phosphate synthase	233	1921	0.98	0.178(I)	8-10-2001	6-03-2001	157
1K4O	-	Dihydrobutanone phosphate synthase	233	1885	1.10	0.177(I)	8-10-2001	6-03-2002	157
1K4P	-	Dihydrobutanone phosphate synthase	233	1858	1.00	0.178(I)	8-10-2001	6-03-2002	157
1K5C	+	Endopolygalacturonase I	335	2964	0.96	0.114	10-10-2001	5-06-2002	158
1K5N	+	Histocompatibility complex HLA-B*27	385	4126	1.09	0.123	11-10-2001	30-10-2002	159
1K6A	+	Xylanase from <i>T. aurantiacus</i>	303	2497	1.14	0.116	15-10-2001	3-07-2002	160
1K6U	+	Cyclic BPTI	58	633	1.00	0.128	17-10-2001	19-12-2001	161
1GO6	-	Balhimycin	12	1187	0.98	0.129	19-10-2001	13-06-2002	162
1K7C	+	Rhamnogalacturonan acetyltransferase	233	2170	1.12	0.104	19-10-2001	28-12-2001	163

(continued)

TABLE I (continued)

PDB code	SF availability	Protein	Residues	Atoms	Resolution (Å)	R-factor	Deposition date	Release date	References ^a
1GOK	+	Xylanase	303	2560	1.14	0.180(I)	22-10-2001	25-10-2001	160
1K8U	-	S100A6 Ca-binding protein	90	826	1.15	0.186(I)	25-10-2001	10-04-2002	164
1KCC	-	Endopolygalacturonase I complex	335	3043	1.00	0.114	8-11-2001	5-06-2002	158
1KCD	+	Endopolygalacturonase I complex	335	2986	1.15	0.135	8-11-2001	5-06-2002	158
1KDV	+	PSCP, inhibitor AcIAF	374	3240	1.10	0.130	13-11-2001	12-12-2001	165
1KDY	+	PSCP, inhibitor AcIPF	374	3166	1.10	0.134	13-11-2001	12-12-2001	165
1KF2	+	Ribonuclease A, pH 5.2	124	1197	1.10	0.103	19-11-2001	19-12-2001	166
1KF3	+	Ribonuclease A, pH 5.9	124	1197	1.05	0.102	19-11-2001	19-12-2001	166
1KF4	+	Ribonuclease A, pH 6.3	124	1178	1.10	0.104	19-11-2001	19-12-2001	166
1KF5	+	Ribonuclease A, pH 7.1	124	1178	1.15	0.106	19-11-2001	19-12-2001	166
1KF7	+	Ribonuclease A, pH 8.0	124	1181	1.15	0.106	19-11-2001	19-12-2001	166
1KF8	+	Ribonuclease A, pH 8.8	124	1184	1.15	0.106	19-11-2001	19-12-2001	166
1IS9	-	Endoglucanase A	363	3582	1.03	0.131	26-11-2001	4-09-2002	167
1KG2	-	DNA glycosylase	225	2107	1.20	0.134	26-11-2001	26-11-2001	168
1KJQ	+	PurT transformylase	782	7041	1.05	0.190(I)	5-12-2001	28-06-2002	169
1GQV	+	Eosinophil-derived neurotoxin	135	1318	0.98	0.116	5-12-2001	8-03-2002	170
1KMS	+	Dihydrofolate reductase complex	186	1948	1.09	0.131	17-12-2001	10-07-2002	171
1KMW	+	Dihydrofolate reductase complex	186	1878	1.05	0.130	17-12-2001	10-07-2002	171
1KNG	+	CemG oxidoreductase	156	1231	1.14	0.118	18-12-2001	17-07-2002	172
1KNL	-	Xylanase 10A domain	130	1207	1.20	0.152	19-12-2001	19-06-2002	173
1KNM	-	Xylanase 10A domain complex	130	1231	1.20	0.126	19-12-2001	19-06-2002	173
1KOU	+	Photoactive yellow protein	125	1084	1.16	0.162	22-12-2001	3-04-2002	174

IKQP	-	NH ₃ -dependent NAD ⁺ synthetase	542	5173	1.03	0.116	7-01-2002	28-06-2002	175
IKTH	+	Kunitz domain of type VI collagen	58	556	0.95	0.136	16-01-2002	6-02-2002	176
IKWF	-	Inverting glycosidase	363	3358	0.94	0.094	29-01-2002	13-03-2002	177
IKWN	+	Thaumatin	207	1745	1.20	0.128	30-01-2002	11-12-2002	178
IKZK	+	HIV protease, JE-2147 complex	198	1797	1.09	0.151	6-02-2002	3-04-2001	179
IITX	+	Chitinase A1, catalytic domain	419	4018	1.10	0.117	13-02-2002	13-03-2002	180
IGVK	+	Porcine pancreatic elastase	244	2190	0.94	0.122	14-02-2002	19-07-2002	181
IGVT	-	Endothiapepsin complex	333	2914	0.98	0.110	27-02-2002	4-07-2002	182
IGVU	-	Endothiapepsin complex	335	2914	0.94	0.116	27-02-2002	4-07-2002	182
IGVV	-	Endothiapepsin complex	333	2924	1.05	0.126	27-02-2002	4-07-2002	182
IGVX	-	Endothiapepsin complex	335	2862	1.00	0.140	27-02-2002	4-07-2002	182
IGVW	-	Endothiapepsin complex	334	2889	1.00	0.128	27-02-2002	4-07-2002	182
IGVX	-	Endothiapepsin complex	335	2862	1.00	0.140	27-02-2002	4-07-2002	182
IL3K	+	Ribonucleoprotein A1	196	1525	1.10	0.155	27-02-2002	17-04-2002	183
IUA	+	HiPIP from <i>T. tepidum</i>	83	732	0.80	0.092	1-03-2002	20-03-2001	184
IGWE	+	Catalase from <i>M. lysodeikticus</i>	503	4920	0.88	0.089	15-03-2002	26-04-2002	185
IL9L	+	Granulysin	74	737	0.92	0.137	25-03-2002	6-11-2002	186
ILC0	+	Biliverdin IX α reductase	294	2371	1.20	0.222(1)	4-04-2002	17-07-2002	187
ILF7	-	Human complement protein C8 γ	182	1470	1.20	0.223(1)	10-04-2002	12-06-2002	188
ILJN	+	Lysozyme from turkey egg	129	1209	1.19	0.104	22-04-2002	5-06-2002	189
IGYO	+	Ditetraheme cytochrome c ₃	218	2286	1.20	0.130	29-04-2002	24-05-2002	190
ILNI	+	Ribonuclease Sa	192	2068	1.00	0.116	3-05-2002	31-07-2002	191
ILOK	-	Aminopeptidase	291	2573	1.20	0.158	6-05-2002	27-11-2002	192
ILQP	+	Fosfomycin resistance protein	270	2567	1.19	0.156	11-05-2002	11-09-2002	193
ILQT	+	FprA oxidoreductase	912	9020	1.05	0.134	13-05-2002	31-07-2002	194

(continued)

TABLE I (continued)

PDB code	SF availability	Protein	Residues	Atoms	Resolution (Å)	R-factor	Deposition date	Release date	References ^a
1LS1	+	NG domain of Ffh protein	295	2514	1.10	0.137	16-05-2002	16-11-2002	195
1LU0	+	Squash trypsin inhibitor	58	544	1.03	0.120	21-05-2002	28-08-2002	196
1M15	+	Arginine kinase	357	3422	1.20	0.125	17-06-2002	4-12-2002	197
1IX9	-	SOD from <i>E. coli</i> mutant	410	4151	0.90	0.107	17-06-2002	17-12-2002	198
1IXB	-	SOD from <i>E. coli</i> mutant	410	4176	0.90	0.111	18-06-2002	18-12-2002	198
1MIN	-	Nitrogenase MoFe protein	4052	37069	1.16	0.123	19-06-2002	11-09-2002	199
1MIQ	+	Tetraheme cytochrome <i>c</i> ⁺ oxidized	91	1103	0.97	0.135	20-06-2002	14-08-2002	200
1M1R	+	Tetraheme cytochrome <i>c</i> ⁺ reduced	91	1119	1.02	0.149	20-06-2002	14-08-2002	200
1M24	+	Trichotoxin A50E	38	249	0.90	0.075	21-06-2002	6-11-2002	201
1M2D	-	Ferredoxin from <i>A. aeolicus</i>	220	1780	1.05	0.137	22-06-2002	18-09-2002	202
1H11	-	Endoglucanase Cel5A complex	303	2937	1.08	0.109	1-07-2002	8-08-2002	120
1M40	+	TEM-1 β -lactamase	263	2594	0.85	0.091	1-07-2002	17-07-2002	203
1M5A	-	Porcine 2Co ²⁺ -insulin	102	1005	1.20	0.185(I)	9-07-2002	14-08-2002	204
1H1N	+	Endoglucanase from <i>T. aurantiacus</i>	610	5697	1.12	0.144	19-07-2002	12-08-2002	205
1M9Z	-	Human TGF- β type II receptor	111	1011	1.05	0.156(I)	30-07-2002	11-09-2002	206
1MC2	-	Phospholipase A2	122	1211	0.85	0.095	5-08-2002	21-08-2002	207
1ME3	-	Cruzinain with covalent inhibitor	215	2176	1.20	0.105	7-08-2002	18-12-2002	208
1ME4	-	Cruzinain with noncovalent inhibitor	215	2111	1.20	0.096	7-08-2002	18-12-2002	208
1H2J	+	Endoglucanase Cel5A complex	303	2807	1.15	0.118	9-08-2002	15-08-2002	120

1MGO	+	Horse LADH mutant	748	6149	1.20	0.188	15-08-2002	13-11-2002	209
1MJ4	-	Sulfite oxidase cytochrome b_5	82	788	1.20	0.119	26-08-2002	12-09-2002	210
1MNZ	+	Glucose isomerase	388	3475	0.99	0.123	6-09-2002	25-09-2002	211
1MOO	+	Carbonic anhydrase II mutant	260	2355	1.05	0.157	9-09-2002	18-09-2002	212
1MUW	+	Glucose isomerase <i>S. olivochromi</i>	386	3479	0.86	0.125	24-09-2002	6-11-2002	213
1N3L	+	t-RNA synthetase cytokine	372	2981	1.18	0.180	28-10-2002	20-11-2002	214
1O73	+	L-Asparaginase, <i>E. chrysanthemi</i>	1308	11243	1.00	0.110	7-11-2002	4-12-2002	215
1N60	-	CO dehydrogenase, complex	2526	22478	1.19	0.143	8-11-2002	18-12-2002	216
1N62	-	CO dehydrogenase, complex	2526	22381	1.09	0.144	8-11-2002	18-12-2002	216
	-	Avian pancreatic polypeptide	36		0.98	0.156(I)		1983	217
	-	Ostreotide	24		1.04	0.095		1995	218
	-	HEW lysozyme triclinic	129	1001	0.98	n/a		1998	219

^aReferences:

1. A. Wlodawer, J. Walter, R. Huber, and L. Sjölín, *J. Mol. Biol.* **180**, 301 (1984).
2. K. D. Watenpaugh, T. N. Margulis, L. C. Sieker, and L. H. Jensen, *J. Mol. Biol.* **122**, 175 (1978). K. D. Watenpaugh, L. C. Sieker, and L. H. Jensen, *J. Mol. Biol.* **131**, 509 (1979). K. D. Watenpaugh, L. C. Sieker, and L. H. Jensen, *J. Mol. Biol.* **138**, 615 (1980).
3. K. D. Watenpaugh, in "Proceedings of the Molecular Dynamics Workshop, 1984" (J. Hermans, ed.), University of North Carolina Press, Chapel Hill, NC, 1985.
4. S. P. Wood, I. J. Tickle, A. M. Treharne, J. E. Pitts, Y. Mascarenhas, J. Y. Li, J. Husain, S. Cooper, T. L. Blundell, V. J. Hruby *et al.*, *Science* **232**, 633 (1986).
5. Z. Dauter, L. C. Sieker, and K. S. Wilson, *Acta Crystallogr. B* **48**, 42 (1992).
6. M. M. Teeter, S. M. Roe, and N. H. Heo, *J. Mol. Biol.* **230**, 292 (1993). B. Stec, R. Zhou, and M. M. Teeter, *Acta Crystallogr. D* **51**, 663 (1995). A. Yamano, N. H. Heo, and M. M. Teeter, *J. Biol. Chem.* **272**, 9597 (1997).
7. G. Scapin, J. I. Gordon, and J. C. Sacchettini, *J. Biol. Chem.* **267**, 4253 (1992).

(continued)

8. B. Zhao, M. Carson, S. E. Ealick, and C. E. Bugg, *J. Mol. Biol.* **227**, 239 (1992).
9. S. Tsunawano, T. Masaki, M. Hirose, M. Soejima, and F. Sakiyama, *J. Biol. Chem.* **264**, 3832 (1989).
10. A. Yamano and M. M. Teeter, *J. Biol. Chem.* **269**, 13956 (1994).
11. T. Sato, H. Ozaki, Y. Hata, Y. Kitagawa, Y. Katsube, and Y. Shimomishi, *Biochemistry* **33**, 8641 (1994).
12. H. Ozaki, T. Sato, H. Kubota, Y. Hata, Y. Katsube, and Y. Shimomishi, *J. Biol. Chem.* **266**, 5934 (1991).
13. J. P. Derrick and D. B. Wigley, *J. Mol. Biol.* **243**, 906 (1994).
14. S. Parkin, B. Rupp, and H. Hope, *Acta Crystallogr. D* **52**, 18 (1996).
15. J. Sevcik, Z. Dauter, V. Lamzin, and K. S. Wilson, *Acta Crystallogr. D* **52**, 327 (1996).
16. C. Frazão, C. M. Soares, M. A. Carrondo, E. Pohl, Z. Dauter, K. S. Wilson, M. Hervás, J. A. Navarro, M. A. De la Rosa, and G. M. Sheldrick, *Structure* **3**, 1159 (1995).
17. L. Tong, T. C. Warren, J. King, R. Betageri, J. Rose, and S. Jakes, *J. Mol. Biol.* **256**, 601 (1996).
18. Z. Dauter, K. S. Wilson, L. C. Sieker, J. M. Moulis, and J. Meyer, *Proc. Natl. Acad. Sci. USA* **93**, 8836 (1996).
19. D. H. Anderson, M. S. Weiss, and D. Eisenberg, *Acta Crystallogr. D* **52**, 469 (1996). D. H. Anderson, M. S. Weiss, and D. Eisenberg, *J. Mol. Biol.* **273**, 479 (1996).
20. S. H. Hu, J. Gehrmann, L. W. Guddat, P. F. Alewood, D. J. Craik, and J. L. Martin, *Structure* **4**, 417 (1996).
21. V. S. Kumaraswamy, P. F. Lindley, C. Slingsby, and I. D. Glover, *Acta Crystallogr. D* **52**, 611 (1996).
22. L. W. Guddat, J. A. Martin, L. Shan, A. B. Edmondson, and W. R. Gay, *Biochemistry* **35**, 11329 (1996).
23. J. R. Tame, S. H. Sleight, A. J. Wilkinson, and J. E. Ladbury, *Nat. Struct. Biol.* **3**, 998 (1996). J. R. H. Tame, *Acta Crystallogr. D* **56**, 1554 (2000).
24. Z. Wang, H. Luecke, N. Yao, and F. A. Quirocho, *Nat. Struct. Biol.* **4**, 519 (1997).
25. S. Parkin, B. Rupp, and H. Hope, *Acta Crystallogr. D* **52**, 1161 (1996).
26. K. Mergeau, B. Arnoux, D. Perahia, K. Norris, F. Norris, and A. Ducruix, *Acta Crystallogr. D* **54**, 306 (1998).
27. P. J. Loll, A. E. Bevivino, B. D. Korty, and P. H. Axelsen, *J. Am. Chem. Soc.* **119**, 1516 (1997).
28. A. Deacon, T. Gleichmann, A. J. Kalb, H. Price, J. Raftery, G. Bradbrook, J. Yarniv, and J. R. Helliwell, *J. Chem. Soc. Faraday Trans. 1997*, 4305 (1997).
29. G. M. Sheldrick, Z. Dauter, K. S. Wilson, H. Hope, and L. C. Sieker, *Acta Crystallogr. D* **49**, 18 (1993).
30. S. Longhi, M. Czjzek, V. Lamzin, A. Nicolas, and C. Cambillau, *J. Mol. Biol.* **268**, 779 (1997).
31. E. Schmitt, Y. Mechulam, M. Fromant, P. Plateau, and S. Blanquet, *EMBO J.* **16**, 4760 (1997).
32. M. A. Walsh, T. R. Schneider, L. C. Sieker, Z. Dauter, V. S. Lamzin, and K. S. Wilson, *Acta Crystallogr. D* **54**, 522 (1998).
33. G. D. Smith, R. H. Blessing, S. E. Ealick, J. C. Fontecilla-Camps, H. A. Hauptman, D. Housset, and D. A. Langa, *Acta Crystallogr. D* **53**, 551 (1997).

34. K. S. Wilson, S. Butterworth, Z. Dauter, V. S. Lamzin, M. Walsh, S. Wodak, S. Pontius, J. Richelle, A. Vaguine, C. Sander, R. W. W. Hoofst, G. Vriend, J. M. Thornton, R. A. Laskowski, M. W. MacArthur, E. J. Dodson, G. Murshudov, T. J. Oldfield, T. Kaptein, and J. A. C. Rullman, *J. Mol. Biol.* **276**, 417 (1998). G. M. Sheldrick, Z. Dauter, K. S. Wilson, H. Hope, and L. C. Sieker, *Acta Crystallogr. D* **49**, 18 (1993).
35. S. H. Hu, J. Gehrman, P. F. Alewood, D. J. Craik, and J. L. Martin, *Biochemistry* **36**, 11323 (1997).
36. B. M. Burkhart, R. M. Gassman, D. A. Langs, W. A. Pangborn, and W. L. Duax, *Biophys. J.* **75**, 2135 (1998).
37. M. Schäfer, T. R. Schneider, and G. M. Sheldrick, *Structure* **4**, 1509 (1996).
38. D. M. Lawson, C. E. Williams, L. A. Mitchenall, and R. N. Pau, *Structure* **6**, 1529 (1998). D. M. Lawson, C. E. Williams, D. J. White, A. P. Choay, L. A. Mitchenall, and R. N. Pau, *J. Chem. Soc. Dalton Trans.* **1997**, 3981 (1997).
39. U. Ermiler, W. Grabarse, S. Shima, M. Goubaud, and R. K. Thauer, *Science* **278**, 1457 (1997).
40. Z. Dauter, K. S. Wilson, L. C. Sieker, J. Meyer, and J. M. Moulis, *Biochemistry* **36**, 16065 (1997).
41. L. K. Steinrauf, *Acta Crystallogr. D* **54**, 767 (1998).
42. M. Vlasi, Z. Dauter, K. S. Wilson, and M. Kokkinidis, *Acta Crystallogr. D* **54**, 1245 (1998).
43. D. E. Brodersen, M. Etzlerodt, P. Madsen, J. E. Celis, H. C. Thogersen, J. Nyborg, and M. Kjeldgaard, *Structure* **6**, 477 (1998).
44. D. E. Brodersen, E. de La Fortelle, C. Vonrhein, G. Bricogne, J. Nyborg, and M. Kjeldgaard, *Acta Crystallogr. D* **56**, 431 (2000).
45. S. H. Hu, M. Loughnan, R. Miller, C. M. Weeks, R. H. Blessing, P. F. Alewood, R. J. Lewis, and J. L. Martin, *Biochemistry* **37**, 11425 (1998).
46. K. Harata, Y. Abe, and M. Muraki, *Proteins* **30**, 232 (1998).
47. J. Vojtechovsky, K. Chu, J. Berendzen, R. M. Sweet, and I. Schlichting, *Biophys. J.* **77**, 2153 (1999).
48. S. Karlens, L. F. Iversen, I. K. Larsen, H. J. Flodgaard, and J. S. Kastrop, *Acta Crystallogr. D* **54**, 598 (1998).
49. M. Schäfer, G. M. Sheldrick, I. Bahner, and H. Lackner, *Angew. Chem. Int. Ed.* **37**, 2381 (1998).
50. E. F. Garman, J. Wouters, T. R. Schneider, E. R. Vimr, W. G. Laver, and G. M. Sheldrick, *Acta Crystallogr. A* **52**, C8 (IUCr Congress 1996 Abstract) (1996).
51. S. Banuelos, M. Saraste, and K. D. Carugo, *Structure* **6**, 1419 (1998).
52. U. K. Genick, S. M. Soltis, P. Kuhn, I. L. Canestrelli, and E. D. Getzoff, *Nature* **392**, 206 (1998).
53. G. J. Davies, L. Mackenzie, A. Varrat, M. Dauter, A. M. Brzozowski, M. Schülein, and S. G. Withers, *Biochemistry* **37**, 11707 (1998).
54. R. Bau, D. C. Rees, D. M. Kurtz, R. A. Scott, H. S. Huang, M. W. W. Adams, and M. K. Eidsness, *J. Biol. Inorg. Chem.* **3**, 484 (1998).
55. Y. Guan, R. C. Manuel, A. S. Arvai, S. S. Parikh, C. D. Mol, J. H. Miller, S. Lloyd, and J. A. Tainer, *Nat. Struct. Biol.* **5**, 1058 (1998).
56. D. Ghosh, M. Erman, M. Sawicki, P. Lala, D. R. Weeks, N. Li, W. Pangborn, D. J. Thiel, H. Jornvall, R. Gutierrez, and J. Eyzaguirre, *Acta Crystallogr. D* **55**, 779 (1999).
57. P. Kuhn, M. Knapp, M. Soltis, G. Ganshaw, M. Thoene, and R. Bott, *Biochemistry* **37**, 13446 (1998).
58. L. Lo Leggio, S. Kalogiannis, M. K. Bhat, and R. W. Pickersgill, *Proteins* **36**, 295 (1999).

(continued)

58. J. P. Declercq, C. Evrard, V. Lamzin, and J. Parello, *Protein Sci.* **8**, 2194 (1999).
59. A. R. Khan, J. C. Parrish, M. E. Fraser, W. W. Smith, P. A. Bartlett, and M. N. James, *Biochemistry* **37**, 16839 (1998).
60. S. Freitag, I. Le Trong, L. A. Klumb, P. S. Stayton, and R. E. Stenkamp, *Acta Crystallogr. D* **55**, 1118 (1999).
61. I. Usón, G. M. Sheldrick, E. de La Fortelle, G. Bricogne, S. Di Marco, J. P. Priestle, M. G. Grütter, and P. R. Mittl, *Structure* **7**, 55 (1999).
62. T. Sandalova, G. Schneider, H. Käck, and Y. Lindqvist, *Acta Crystallogr. D* **55**, 610 (1999).
63. G. G. Prive, D. H. Anderson, L. Wesson, D. Cascio, and D. Eisenberg, *Protein Sci.* **8**, 1400 (1999).
64. W. R. Patterson, D. H. Anderson, W. F. Degrado, D. Cascio, and D. Eisenberg, *Protein Sci.* **8**, 1410 (1999).
65. G. S. Kachalova, A. N. Popov, and H. D. Bartunik, *Science* **284**, 473 (1999).
66. G. G. Sulzenbacher, L. F. Mackenzie, K. S. Wilson, S. G. Withers, C. Dupont, and G. J. Davies, *Biochemistry* **38**, 4826 (1999).
67. E. Parisini, F. Capozzi, P. Lubini, V. Lamzin, C. Luchinat, and G. M. Sheldrick, *Acta Crystallogr. D* **55**, 1773 (1999).
68. I. S. Ridder, H. J. Rozeboom, and B. W. Dijkstra, *Acta Crystallogr. D* **55**, 1273 (1999).
69. A. Varrot, M. Schülein, M. Pipeliet, A. Vasella, and G. J. Davies, *J. Am. Chem. Soc.* **121**, 2621 (1999).
70. K. Harata, Y. Abe, and M. Muraki, *J. Mol. Biol.* **287**, 347 (1999).
71. A. C. Rosenzweig, D. L. Huffman, M. Y. Hou, A. K. Wermimont, R. A. Pufahl, and T. V. O'Halloran, *Structure* **7**, 605 (1999).
72. M. Ferraroni, W. Rypniewski, K. S. Wilson, M. S. Viezzoli, L. Banci, I. Bertini, and S. Mangani, *J. Mol. Biol.* **288**, 413 (1999).
73. S. Misra and J. H. Hurley, *Cell* **97**, 657 (1999).
74. M. Weik, R. B. G. Ravelli, G. Kryger, S. McSweeney, M. L. Raves, M. Harel, P. Gros, I. Silman, J. Kroon, and J. L. Sussman, *Proc. Natl. Acad. Sci. USA* **97**, 623 (2000).
75. A. I. Derman, T. Mau, and D. A. Agard, in preparation.
76. K. Gruber, M. Guggamig, U. G. Wagner, and C. Kratky, *Biol. Chem.* **380**, 993 (1999).
77. W. Watt, K. A. Koepflinger, A. M. Mildner, R. L. Heinrichson, A. G. Tomasselli, and K. D. Watenpaugh, *Structure* **7**, 11358 (1999).
78. D. J. Hosfield, Y. Guan, B. J. Haas, R. P. Cunningham, and J. A. Tainer, *Cell* **98**, 397 (1999).
79. K. Volz, *Protein Sci.* **8**, 2428 (1999).
80. S. Umhau, L. Pollegioni, G. Molla, K. Diederichs, W. Welte, S. M. Pilone, and S. Ghisla, *Proc. Natl. Acad. Sci. USA* **97**, 12463 (2000).
81. U. Mueller, D. Peri, F. X. Schmid, and U. Heinemann, *J. Mol. Biol.* **297**, 975 (2000).
82. S. V. Shlyapnikov, V. V. Lunin, M. Perbandt, K. M. Polyakov, V. Y. Lunin, V. M. Levdikov, C. Betzel, and A. M. Mikhailov, *Acta Crystallogr. D* **56**, 567 (2000).
83. J. Lubkowski, Z. Dauter, F. Yang, J. Alexandratos, G. Merkel, A. M. Skalka, and A. Wlodawer, *Biochemistry* **38**, 13512 (1999).
84. R. Morales, M. H. Charon, G. Hudry-Clergeon, Y. Pettillot, S. Norager, M. Medina, and M. Frey, *Biochemistry* **38**, 15764 (2000).
85. P. C. Bourne, M. N. Isupov, and J. A. Littlechild, *Structure* **8**, 143 (2000).
86. S. A. Roberts, A. Weichsel, Y. Qiu, J. A. Shelnett, F. A. Walker, and W. R. Montfort, *Biochemistry* **40**, 11327 (2001).
87. F. Poy, M. B. Yaffe, J. Sayos, K. Saxena, M. Morra, J. Sumegi, L. C. Cantley, C. Terhorst, and M. J. Eck, *Mol. Cell* **4**, 555 (1999).

88. P. Luan, A. Heine, K. Zeng, B. Moyer, S. E. Greasley, P. Kuhn, W. E. Balch, and I. A. Wilson, *Traffic* **1**, 270 (2000).
89. M. Württele, M. Hahn, K. Hilpert, and W. Höhne, *Acta Crystallogr. D* **56**, 520 (2000).
90. T. R. Schneider, J. Körcher, E. Pohl, P. Lubini, and G. M. Sheldrick, *Acta Crystallogr. D* **56**, 705 (2000).
91. F. Yang, P. Forrer, Z. Dauter, J. F. Conway, N. Cheng, M. E. Cerritelli, A. C. Steven, A. Plückthun, and A. Wlodawer, *Nat. Struct. Biol.* **7**, 230 (2000).
92. R. J. Guan, X. L. Xe, M. Wang, Ye-Xiang, G. P. Li, D. C. Wang, *Protein Pept. Lett.* **9**, 441 (2002).
93. Z. H. Zhang, J. Ren, D. K. Stammers, J. E. Baldwin, K. Harlos, and C. J. Schofield, *Nat. Struct. Biol.* **7**, 127 (2000).
94. L. Esposito, L. Vitagliano, F. Sica, G. Sorrentino, A. Zagari, and L. Mazzarella, *J. Mol. Biol.* **297**, 713 (2000). L. Esposito, G. Vitagliano, A. Zagari, and L. Mazzarella, *Protein Eng.* **13**, 825 (2000).
95. S. Benini, A. Gonzalez, W. R. Rypniewski, K. S. Wilson, J. J. van Beeumen, and S. Ciurli, *Biochemistry* **39**, 13115 (2000).
96. G. Kurisu, Y. Kai, and S. Harada, *J. Inorg. Biochem.* **82**, 225 (2000).
97. C. Jelsch, M. M. Teeter, V. Lamzin, V. Pichon-Lesme, B. Blessing, and C. Lecomte, *Proc. Natl. Acad. Sci. USA* **97**, 3171 (2000). V. S. Lamzin, R. J. Morris, Z. Dauter, K. S. Wilson, and M. M. Teeter, *J. Biol. Chem.* **274**, 20753 (1999).
98. V. Ducros, S. J. Charnock, U. Derewenda, Z. S. Derewenda, Z. Dauter, C. Dupont, F. Shareck, R. Morosoli, D. Kluepfel, and G. J. Davies, *J. Biol. Chem.* **275**, 23020 (2000).
99. L. Jin, S. L. Briggs, S. Chandrasekhar, N. Y. Chirgadze, D. K. Clawson, R. W. Schevitz, D. L. Smiley, A. H. Tashjian, and F. Zhang, *J. Biol. Chem.* **275**, 27238 (2000).
100. A. Gonzalez, E. Cedergren, G. Larsson, and R. Persson, *Acta Crystallogr. D* **57**, 767 (2001).
101. M. A. Wilson and A. T. Brunger, *J. Mol. Biol.* **301**, 1237 (2000).
102. M. P. Sebastiao, V. Lamzin, M. J. Saraiva, and A. M. Damas, *J. Mol. Biol.* **306**, 733 (2001).
103. P. Kuhn, A. M. Deacon, S. Comoso, G. Rajaseger, R. M. Kini, I. Usón, and P. R. Kolatkar, *Acta Crystallogr. D* **56**, 1401 (2000).
104. R. Brudler, T. E. Meyer, U. K. Genick, S. Devanathan, T. T. Woo, D. P. Millar, K. Gerwert, M. A. Cusanovich, G. Tollin, and E. D. Getzoff, *Biochemistry* **39**, 13478 (2000). U. K. Genick, S. M. Soltis, P. Kuhn, I. L. Castrenelli, and E. D. Getzoff, *Nature* **392**, 206 (1998).
105. W. P. Burmeister, S. Cottaz, P. Rollin, A. Vasella, and B. Henrissat, *J. Biol. Chem.* **275**, 39385 (2000).
106. R. Meijers, R. J. Morris, H. W. Adolph, A. Merli, V. S. Lamzin, and E. S. Cedergren-Zepezauer, *J. Biol. Chem.* **276**, 9316 (2001).
107. A. Heroux, E. L. White, L. J. Ross, A. P. Kuzin, and D. W. Borhani, *Structure* **8**, 1309 (2000).
108. W. R. Rypniewski, P. R. Ostergaard, M. Norregaard-Madsen, M. Dauter, and K. S. Wilson, *Acta Crystallogr. D* **57**, 8 (2001).
109. S. Teixeira, L. Lo Leggio, R. Pickersgill, and C. Cardin, *Acta Crystallogr. D* **57**, 385 (2001).
110. K. Hakansson, A. H. J. Wang, and C. G. Miller, *Proc. Natl. Acad. Sci. USA* **97**, 14097 (2000).
111. P. Heikinheimo, V. Tuominen, A.-K. Ahonen, A. Tepliyakov, B. S. Cooperman, A. A. Baykov, R. Lahti, and A. Goldman, *Proc. Natl. Acad. Sci. USA* **98**, 3121 (2001).

(continued)

112. R. Bertisio, A. Viguera, L. Serrano, and M. Wilmanns, *Acta Crystallogr. D* **57**, 337 (2001).
113. R. B. Rose, J. A. Endrizzzi, J. D. Cronk, J. Holton, and T. Alber, *Biochemistry* **39**, 15062 (2000).
114. C. S. Bond, M. P. Shaw, M. S. Alphey, and W. N. Hunter, *Acta Crystallogr. D* **57**, 755 (2001).
115. R. A. Steiner, H. J. Rozeboom, A. de Vries, K. H. Kalk, G. N. Murshudov, K. S. Wilson, and B. W. Dijkstra, *Acta Crystallogr. D* **57**, 516 (2001).
116. D. Ghosh, M. Sawicki, P. Lala, M. Erman, W. Pangborn, J. Eyzaguirre, R. Gutierrez, H. Jornvall, and D. J. Thiel, *J. Biol. Chem.* **276**, 11159 (2001).
117. A. Adlagatta, S. Krzywda, H. Czapińska, J. Otlewski, and M. Jaskólski, *Acta Crystallogr. D* **57**, 649 (2001).
118. G. D. Smith, W. A. Pangborn, and R. H. Blessing, *Acta Crystallogr. D* **57**, 1091 (2001).
119. P. J. B. Pereira, S. Macedo-Ribeiro, A. Parraga, R. Perez-Luque, O. Cunningham, K. Darcy, T. J. Mantle, and M. Coll, *Nat. Struct. Biol.* **8**, 215 (2001).
120. A. Varrot and G. J. Davies, *Acta Crystallogr. D* **59**, 447 (2003).
121. A. Wlodawer, M. Li, Z. Dauter, A. Gustchina, K. Uchida, and K. Oda, *Nat. Struct. Biol.* **8**, 442 (2001).
122. A. A. Antson, D. J. Smith, D. I. Roper, S. Lewis, L. S. D. Caves, C. S. Verma, S. L. Buckley, P. J. Lillford, and R. E. Hubbard, *J. Mol. Biol.* **305**, 875 (2001).
123. P. Burkhard, M. Meier, and A. Lustig, *Protein Sci.* **9**, 2294 (2000).
124. W. Lee, M. A. McDonough, L. P. Kotra, Z.-H. Li, N. R. Silvaggi, Y. Takeda, J. A. Kelly, and S. Mobashery, *Proc. Natl. Acad. Sci. USA* **98**, 1427 (2001).
125. H.-K. Schroder Leiros, S. M. McSweeney, and A. O. Smålas, *Acta Crystallogr. D* **57**, 488 (2001).
126. K. Kamada, J. De Angelis, R. G. Roeder, and S. K. Burley, *Proc. Natl. Acad. Sci. USA* **98**, 3115 (2001).
127. R. C. Deo, N. Somenberg, and S. K. Burley, *Proc. Natl. Acad. Sci. USA* **98**, 4414 (2001).
128. D. A. R. Sanders, D. N. Moothoo, J. Raftery, A. J. Howard, J. R. Helliwell, and J. H. Naismith, *J. Mol. Biol.* **310**, 875 (2001).
129. V. R. Samyгина, A. N. Popov, E. V. Rodina, N. N. Vorobyeva, V. S. Lamzin, K. M. Polyakov, S. A. Kurilova, T. I. Nazarova, and S. M. Avaeva, *J. Mol. Biol.* **314**, 633 (2001).
130. E. J. Gordon, G. A. Leonard, S. McSweeney, and P. F. Zagalsky, *Acta Crystallogr. D* **57**, 1230 (2001).
131. A. Pesce, S. Dewilde, L. Kiger, M. Milani, P. Ascenzi, M. C. Marden, M. L. Van Hauwaert, J. Vanfleteren, L. Moens, and M. Bolognesi, *J. Mol. Biol.* **309**, 1153 (2001).
132. E. Gavuzzo, G. Pochetti, F. Mazza, C. Gallina, B. Gorini, S. D'Alessio, M. Pieper, H. Tschesche, and P. A. Tucker, *J. Med. Chem.* **43**, 3377 (2000).
133. S. Geremia, G. Garau, R. Sgarra, M. S. Viezzoli, and L. Randaccio, *Protein Sci.* **11**, 6 (2002).
134. C. Betzel, S. Gourinath, P. Kumar, P. Kaur, M. Perbandt, S. Eschenburg, and T. P. Singh, *Biochemistry* **40**, 3080 (2001).
135. C. Sauter, F. Otalora, J.-A. Gavira, O. Vidal, R. Giegé, and J.-M. Garcia-Ruiz, *Acta Crystallogr. D* **57**, 1119 (2001).

136. W. Grabarse, F. Mahlert, E. C. Duin, M. Goubeaud, S. Shima, R. K. Thauer, V. Lamzin, and U. Ermler, *J. Mol. Biol.* **309**, 315 (2001).
137. D. M. Hoover, O. Chertov, and J. Lubkowsky, *J. Biol. Chem.* **276**, 39021 (2001).
138. E. Sabini, S. Danielsen, M. Schülein, G. J. Davies, and K. S. Wilson, *Acta Crystallogr. D* **57**, 1344 (2001).
139. P. R. Seavers, R. J. Lewis, J. A. Brannigan, K. H. Verschueren, G. N. Murshudov, and A. J. Wilkinson, *Structure* **9**, 605 (2001).
140. S. N. Ruzheinikov, S. K. Das, S. E. Sedelnikova, A. Hartley, S. J. Foster, M. J. Horsburgh, A. G. Cox, C. W. McCleod, A. Mekhalifa, A. M. Blackburn, D. W. Rice, and P. J. Baker, *J. Mol. Biol.* **313**, 111 (2001).
141. A. Varrot, M. Schülein, S. Fruchard, H. Driguez, and G. J. Davies, *Acta Crystallogr. D* **57**, 1739 (2001).
142. M. Simonovic and K. Volz, *J. Biol. Chem.* **276**, 28637 (2001).
143. A. Heine, G. DeSantis, J. G. Luz, M. Mitchell, C.-H. Wong, and I. A. Wilson, *Science* **294**, 369 (2001).
144. O. Almog, A. Gonzalez, D. Klein, S. Braun, and G. Shoham, in preparation.
145. H. Bönisch, C. L. Schmidt, G. Schäfer, and R. Ladenstein, *J. Mol. Biol.* **319**, 791 (2002).
146. I. Zegers, J. C. Martins, R. Willem, L. Wyns, and J. Messens, *Nat. Struct. Biol.* **8**, 843 (2001).
147. H. Shimizu, S.-Y. Park, Y. Shiro, and S.-I. Adachi, *Acta Crystallogr. D* **58**, 81089 (2002).
148. S. C. Griffith, M. R. Sawaya, D. R. Boutz, N. Thapar, J. E. Katz, S. Clarke, and T. O. Yates, *J. Mol. Biol.* **313**, 1103 (2001).
149. R. Lang, A. Kocourek, M. Braun, H. Tschesche, R. Huber, W. Bode, and K. Maskos, *J. Mol. Biol.* **312**, 731 (2001).
150. K. E. McAuley, Y. Jia-Xing, E. J. Dodson, J. Lehmebeck, P. R. Oestergaard, and K. S. Wilson, *Acta Crystallogr. D* **57**, 1573 (2001).
151. M. Baedeker and G. Schultz, *Eur. J. Biochem.* **269**, 1790 (2002).
152. K. Fukuyama, T. Okada, Y. Kakuta, and Y. Takahashi, *J. Mol. Biol.* **315**, 1155 (2002).
153. M. M. Teeter, A. Yamano, B. Stec, and U. Mohanty, *Proc. Natl. Acad. Sci. USA* **98**, 11242 (2001).
154. A. Spallarosa, J. L. Donahue, T. J. Larson, M. Bolognesi, and D. Bordo, *Structure* **9**, 1117 (2001).
155. B. Mahalingam, P. Boross, Y.-F. Wang, J. M. Louis, C. C. Fischer, J. Tozser, R. W. Harrison, and I. T. Weber, *Proteins Struct. Funct. Genet.* **48**, 107 (2002).
156. C. Chang, D. L. Newton, S. M. Rybak, and A. Wlodawer, *J. Mol. Biol.* **317**, 119 (2002).
157. D.-I. Liao, Y.-J. Zheng, P. V. Viitanen, and D. B. Jordan, *Biochemistry* **41**, 1795 (2002).
158. T. Shimizu, T. Nakatsu, K. Miyairi, T. Okuno, and H. Kato, *Biochemistry* **41**, 6651 (2002).
159. M. Hülsmeier, R. C. Hillig, A. Volz, M. Rühl, W. Schröder, W. Saenger, A. Ziegler, and B. Uchanska-Ziegler, *J. Biol. Chem.* **49**, 47844 (2002).
160. L. Lo Leggio, S. Kalogiannis, K. Eckert, S. C. M. Teixeira, M. K. Bhat, C. Andrei, R. W. Pickersgill, and S. Larsen, *FEBS Lett.* **509**, 303 (2002).
161. I. Botos, Z. Wu, W. Lu, and A. Wlodawer, *FEBS Lett.* **509**, 90 (2001).
162. C. Lehmann, G. Bunkóczi, L. Vértessy, and G. M. Sheldrick, *J. Mol. Biol.* **318**, 723 (2002).
163. A. Molgaard and S. Larsen, *Acta Crystallogr. D* **58**, 111 (2002).

(continued)

164. L. R. Oltterbein, J. Kordowska, C. Witte-Hoffmann, C.-L. A. Wang, and R. Dominguez, *Structure* **10**, 557 (2002).
165. A. Wlodawer, M. Li, A. Gustchina, Z. Dauter, K. Uchida, H. Oyama, N. E. Goldfarb, B. M. Dunn, and K. Oda, *Biochemistry* **40**, 15602 (2001).
166. R. Berisio, V. S. Lamzin, F. Sica, K. S. Wilson, A. Zagari, and L. Mazzarella, *J. Mol. Biol.* **292**, 845 (1999). R. Berisio, F. Sica, V. S. Lamzin, K. S. Wilson, A. Zagari, and L. Mazzarella, *Acta Crystallogr. D* **58**, 441 (2002).
167. A. Schmidt, A. Gonzalez, R. J. Morris, M. Costabel, P. M. Alzari, and V. S. Lamzin, *Acta Crystallogr. D* **58**, 1433 (2002).
168. R. Gilboa, D. O. Zharkov, G. Golan, A. S. Fernandes, S. E. Gerchman, E. Matz, J. H. Kycia, A. P. Grollman, and G. Shoham, *J. Biol. Chem.* **277**, 19811 (2002).
169. J. B. Thoden, S. M. Firestine, S. J. Benkovic, and H. M. Holden, *J. Biol. Chem.* **277**, 23898 (2002).
170. G. J. Swaminathan, D. E. Holloway, K. Veluraja, and K. R. Acharya, *Biochemistry* **41**, 3341 (2002).
171. A. E. Klon, A. Héroux, L. J. Ross, V. Pathak, C. A. Johnson, J. R. Piper, and D. W. Borhani, *J. Mol. Biol.* **320**, 677 (2002).
172. M. A. Edeling, L. W. Guddat, R. A. Fabianek, L. Thöny-Meyer, and J. L. Martin, *Structure* **10**, 973 (2002).
173. V. Notenboom, A. B. Boraston, S. J. Williams, D. G. Kilburn, and D. R. Rose, *Biochemistry* **41**, 4246 (2002).
174. D. M. F. van Aalten, W. Crielgaard, K. J. Hellingwerf, and L. Joshua-Tor, *Acta Crystallogr. D* **58**, 585 (2002).
175. J. Symersky, Y. Devedjiev, K. Moore, C. Brouillette, and L. DeLuca, *Acta Crystallogr. D* **58**, 1138 (2002).
176. B. Arnoux, A. Ducruix, and T. Prangé, *Acta Crystallogr. D* **58**, 1252 (2002).
177. D. M. A. Guérin, M.-B. Lascombe, M. Costabel, H. Souchon, V. Lamzin, P. Béguin, and P. M. Alzari, *J. Mol. Biol.* **316**, 1061 (2002).
178. C. Sauter, B. Lorber, and R. Giegé, *Proteins Struct. Funct. Genet.* **48**, 146 (2002).
179. K. K. Reiling, N. F. Endres, D. S. Dauber, C. S. Craik, and R. M. Stroud, *Biochemistry* **41**, 4582 (2002).
180. T. Matsumoto, T. Nonaka, M. Hashimoto, T. Watanabe, and Y. Mitsui, *Proc. Jpn. Acad.* **B75**, 269 (1999).
181. G. Katona, R. C. Wilmouth, P. A. Wright, G. I. Berglund, J. Hajdu, R. Neutze, and C. J. Schofield, *J. Biol. Chem.* **277**, 21970 (2002).
182. L. Coates, P. T. Erskine, M. P. Crump, S. P. Wood, and J. B. Cooper, *J. Mol. Biol.* **318**, 1405 (2002).
183. J. Vitali, J. Ding, J. Jiang, Y. Zhang, A. R. Krainer, and R.-M. Xu, *Nucleic Acids Res.* **30**, 1531 (2002).
184. L. Liu, T. Nogi, M. Kobayashi, T. Nozawa, and K. Miki, *Acta Crystallogr. D* **58**, 1085 (2002).
185. G. N. Murshudov, A. I. Grebenko, J. A. Brannigan, A. A. Antson, V. V. Barynin, G. G. Dodson, Z. Dauter, K. S. Wilson, and W. R. Melik-Adamyan, *Acta Crystallogr. D* **58**, 1972 (2002).
186. D. H. Anderson, M. R. Sawaya, D. Cascio, W. Ernst, R. Modlin, A. Krensky, and D. Eisenberg, *J. Mol. Biol.* **325**, 355 (2003).
187. F. G. Whitby, J. D. Phillips, C. P. Hill, W. McCoubrey, and M. D. Mames, *J. Mol. Biol.* **319**, 1199 (2002).
188. E. Ortlund, C. L. Parker, S. F. Schreck, S. Gimell, W. Minor, J. M. Sodez, and L. Lebioda, *Biochemistry* **41**, 7030 (2002).
189. K. Harata and R. Kanai, *Proteins Struct. Funct. Genet.* **48**, 53 (2002).
190. D. Aragao, C. Frazao, L. Sieker, G. M. Sheldrick, J. LeGall, and M. A. Carrondo, *Acta Crystallogr. D* **59**, 644 (2003).
191. J. Seveik, V. S. Lamzin, Z. Dauter, and K. S. Wilson, *Acta Crystallogr. D* **58**, 1307 (2002).

192. W. T. Desmarais, D. L. Bienvenue, K. P. Bzymek, R. C. Holz, G. A. Petsko, and D. Ringe, *Structure* **10**, 1063 (2002).
193. C. L. Rife, R. E. Pharris, M. E. Newcomer, and R. N. Armstrong, *J. Am. Chem. Soc.* **124**, 11001 (2002).
194. R. T. Bossi, A. Aliverti, D. Raimondi, F. Fischer, G. Zanetti, D. Ferrari, N. Tahallah, C. S. Maier, A. J. R. Heck, M. Rizzi, and A. Mattevi, *Biochemistry* **41**, 8807 (2002).
195. U. D. Ramirez, G. Minasov, P. J. Focia, R. M. Stroud, P. Walter, P. Kuhn, and D. M. Freymann, *J. Mol. Biol.* **320**, 783 (2002).
196. R. Thaimattam, E. Tykarska, A. Bierzynski, G. M. Sheldrick, and M. Jaskólski, *Acta Crystallogr. D* **58**, 1448 (2002).
197. M. S. Yousef, F. Fabiola, J. L. Gattis, T. Somasundaram, and M. S. Chapman, *Acta Crystallogr. D* **58**, 2009 (2002).
198. B. F. Anderson, R. A. Edwards, M. M. Whittaker, J. W. Whittaker, E. N. Baker, and G. B. Jameson, in preparation.
199. O. Einsle, A. Tezcan, S. L. A. Andrade, B. Schmid, M. Yoshida, J. B. Howard, and D. C. Rees, *Science* **197**, 1696 (2002).
200. D. Leys, T. E. Meyer, A. S. Tsapin, K. H. Nealon, M. A. Cusanovich, and J. J. van Beeumen, *J. Biol. Chem.* **277**, 35703 (2002).
201. J. K. Chugh, H. Brückner, and B. A. Wallace, *Biochemistry* **41**, 12934 (2002).
202. A. P. Yeh, X. I. Ambroggio, S. L. A. Andrade, O. Einsle, C. Chatelet, J. Meyer, and D. C. Rees, *J. Biol. Chem.* **277**, 34499 (2002).
203. G. Minasov, X. Wang, and B. K. Shoichet, *J. Am. Chem. Soc.* **124**, 5333 (2002).
204. J. M. Nicholson, L. C. Perkins, and F. C. Korber, in preparation.
205. F. van Petegem, I. Vandenberghe, M. K. Bhat, and J. van Beeumen, *Biochem. Biophys. Res. Commun.* **296**, 161 (2002).
206. C. C. Boesen, S. Radaev, S. A. Motyka, A. Patamawenu, and P. D. Sun, *Structure* **10**, 913 (2002).
207. Q. Liu, Q. Huang, R. Zhang, C. M. Weeks, C. Jelsch, M. Teng, and L. Niu, *J. Biol. Chem.*, in press.
208. L. Huang, L. S. Brinen, and J. A. Ellman, *Bioorg. Med. Chem.* **11**, 21 (2003).
209. J. K. Rubach and B. V. Plapp, *Biochemistry* **42**, 2907 (2003).
210. M. J. Rudolph, J. L. Johnson, K. V. Rajagopalan, and C. Kisker, *Acta Crystallogr. D* **59**, 1183 (2003).
211. E. Nowak, S. Panjikar, and P. A. Tucker, in preparation.
212. D. Duda, G. Lakshmanan, M. Agbandje-McKenna, C. Tu, D. N. Silverman, and R. McKenna, *Acta Crystallogr. D* **59**, 93 (2003).
213. T. D. Fenn, D. Ringe, and G. A. Petsko, in preparation.
214. X.-L. Yang, R. J. Skene, D. E. McRee, and P. Schimmel, *Proc. Natl. Acad. Sci. USA* **99**, 15369 (2002).
215. J. Lubkowski, M. Dauter, K. Aghaiypour, A. Wlodawer, and Z. Dauter, *Acta Crystallogr. D* **59**, 84 (2003).
216. H. Dobbek, L. Gremer, R. Kiefersauer, R. Huber, and O. Meyer, *Proc. Natl. Acad. Sci. USA* **99**, 15971 (2002).
217. I. Glover, I. Haneef, J. Pitts, S. Wood, D. Moss, I. Tickle, and T. Blundell, *Biopolymers* **22**, 293 (1983).
218. E. Pohl, A. Heine, G. M. Sheldrick, Z. Dauter, T. R. Schneider, K. S. Wilson, and J. Kallen, *Acta Crystallogr. D* **51**, 60 (1995).
219. A. M. Deacon, C. M. Weeks, R. Miller, and S. E. Ealick, *Proc. Natl. Acad. Sci. USA* **95**, 9284 (1998).

Even at atomic resolution it is not possible to model some side chains, which completely lack any interpretable electron density, if they are positioned in very flexible protein regions. Sometimes also N- or C-termini or very flexible loops completely lack electron density and cannot be modeled at all. In such cases it is possible either to omit these parts from the model and represent longer side chains by Gly or Ala, which causes conflicts during structure deposition, or model them in “empty” space with zero occupancy, which can confuse the less “crystallographically knowledgeable” users of the PDB. Such cases should be explicitly listed in the publication and remarked in the deposited file.

Multiple conformations also can be observed for the protein main chain. Sometimes an individual “peptide flip” occurs, wherein an individual peptide plane is rotated. A partial occupancy peptide flip will produce extended disorder over neighboring residues, including their side chains (Fig. 2). The whole stretches of main chain, encompassing several residues, also can follow two alternative pathways.¹⁹ Often the inspection of the

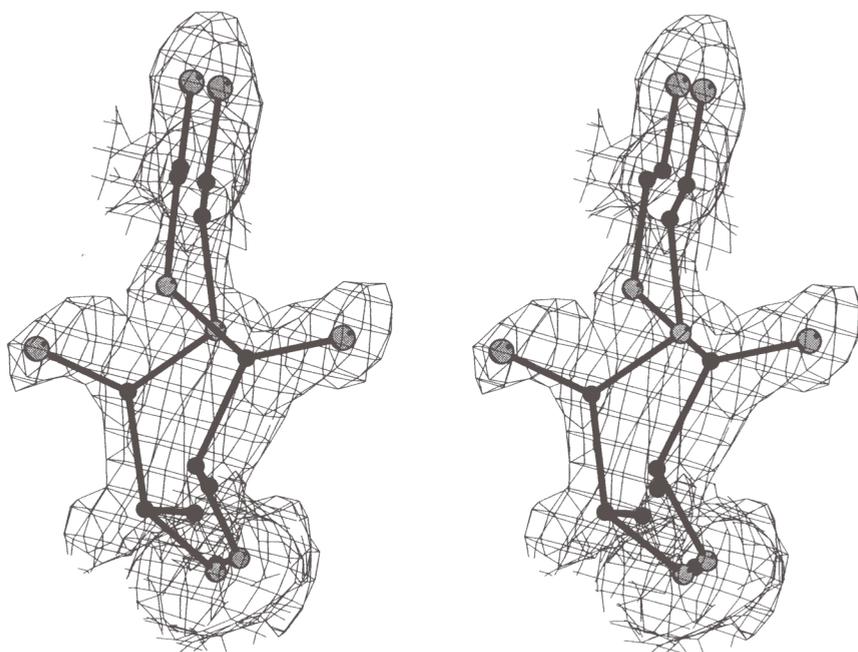


FIG. 2. The $2F_o - F_c$ electron density at the 1σ level of the double conformation peptide-flip region around the Gly D212 residue in the structure of *Erwinia* asparaginase (1O7J), with four molecules of 327 residues each, refined at 1.0 Å.

anisotropic displacement parameters (ADPs) can help in the identification of such split patches of the protein chain.

Most commonly double conformations of side chains occur in polar "linear" amino acids, such as Ser, Asp, Asn, Glu, Gln, Lys, Arg. However, multiple conformations are observed for other residues as well (Fig. 3). It is not uncommon to see the two overlapped rotamers of Val or Thr, differing in χ_1 by 120° and having one γ -site in common. Similarly a Leu side chain may have two χ_2 values 180° apart, with the opposite "chirality" of its C^γ atom. It is relatively common to see a Pro with two alternative C^γ sites, corresponding to the opposite ring pucker. A very clear two orthogonal orientations of the His imidazole ring are apparent in all four independent molecules in the structure of the ASV integrase (Fig. 4).²⁰

In one instance atomic resolution helped clarify the fake double conformation in RNase Sa.²¹ On the basis of the then available protein sequence, a cysteine was modeled in two partially occupied conformations in the structure at 1.8 Å.²² At atomic resolution this residue was reinterpreted as a threonine, having clearly different bond lengths, electron density, and H-bonding contacts of its two γ atoms (Fig. 5).

The double conformations of certain side chains can have a specific relevance for the interpretation of protein biochemistry. The flexibility of certain regions, particularly near the substrate or effector binding sites, may be required for biological activity. Often in crystalline complexes various substrates or inhibitors are not fully occupied, which may result in multiple conformations of those residues, which have to rearrange between the apo and complexed forms. The detailed analysis of such variations, possible only at a very high resolution, may have profound biological importance.

Hydrogen Atoms and H Bonds

X-radiation is scattered by electrons, which are concentrated around atoms in all molecules. Hydrogen atoms, having only one electron, scatter very weakly and present problems even in small-structure crystallography. In view of this, the term "proton," often used by chemists to signify a hydrogen atom, has been deliberately avoided in the present text. According to the traditional opinion, it is not possible to identify hydrogen atoms in macromolecules by X-ray diffraction, since their signal tends to disappear in the noise. However, in view of the results of many

¹⁹ Z. Dauter, K. S. Wilson, L. C. Sieker, J. Meyer, and J. M. Moulis, *Biochemistry* **36**, 16065 (1997).

²⁰ J. Lubkowski, Z. Dauter, F. Yang, J. Alexandratos, G. Merkel, A. M. Skalka, and A. Wlodawer, *Biochemistry* **38**, 13512 (1999).

²¹ J. Sevcik, Z. Dauter, V. Lamzin, and K. S. Wilson, *Acta Crystallogr. D* **52**, 327 (1996).

²² J. Sevcik, E. J. Dodson, and G. G. Dodson, *Acta Crystallogr. B* **47**, 240 (1991).

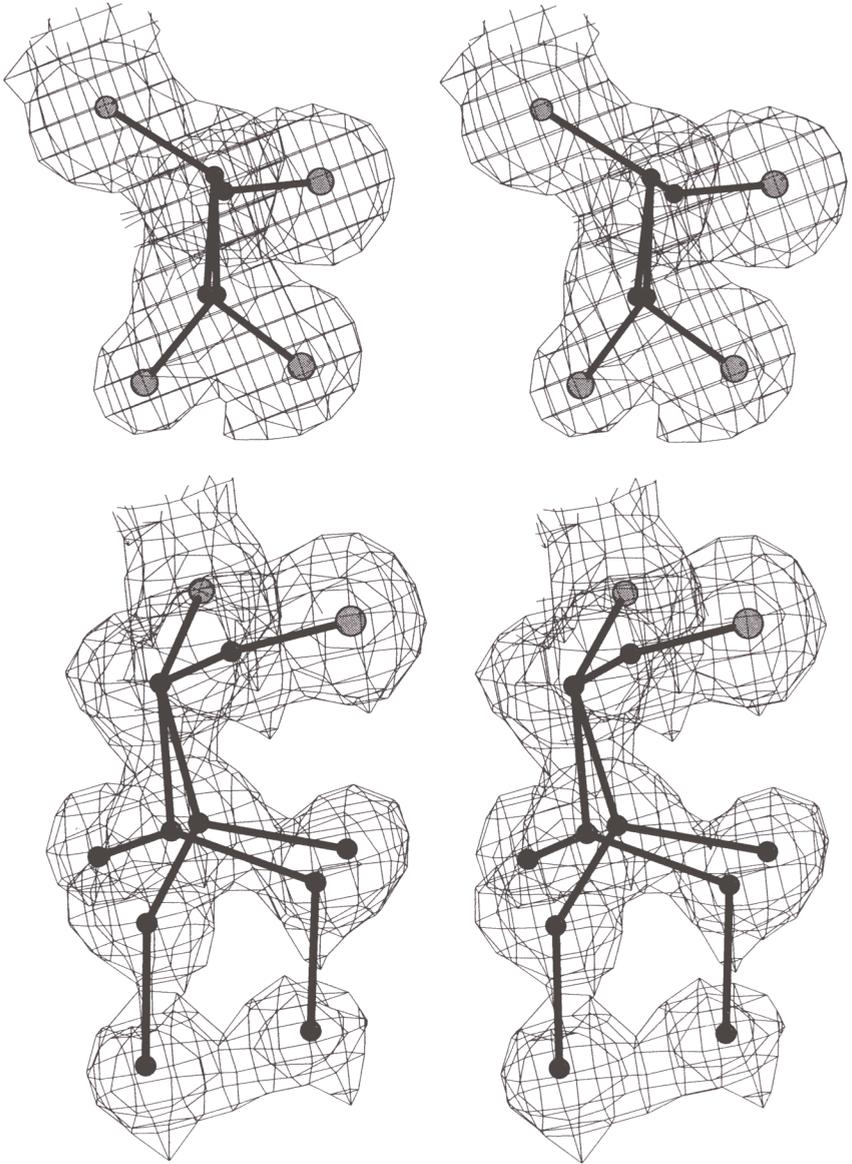


FIG. 3. (continued).

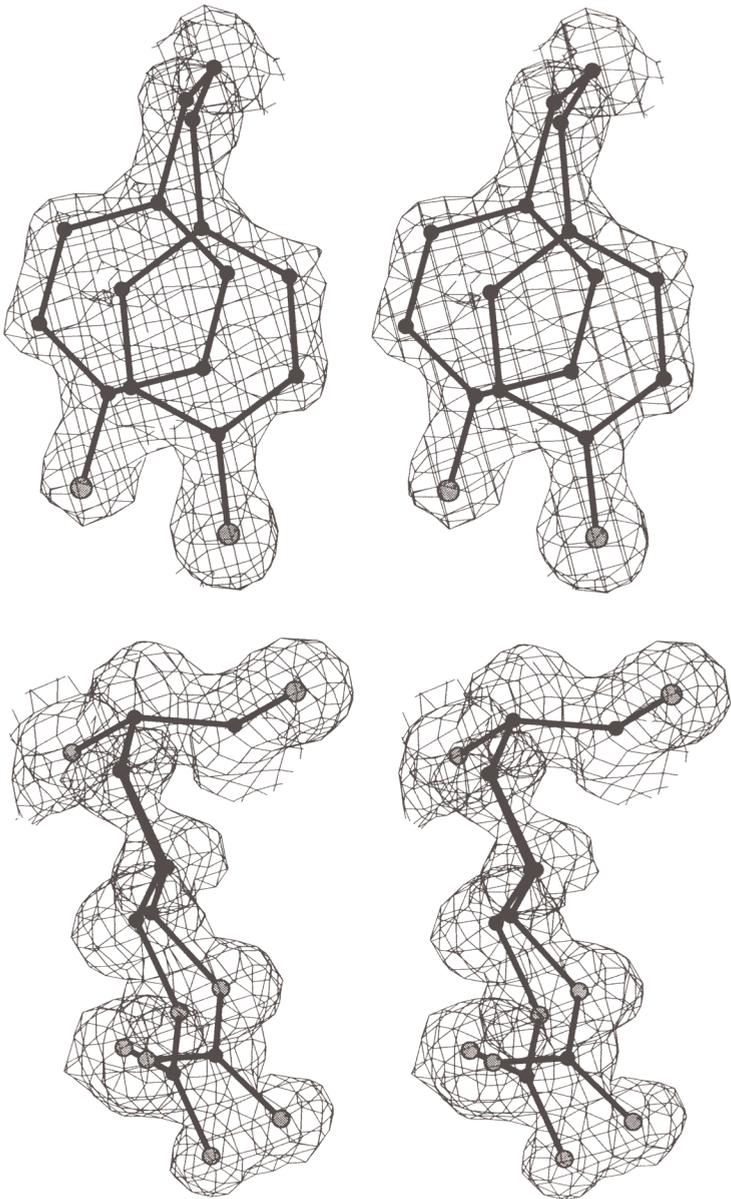


FIG. 3. (continued).

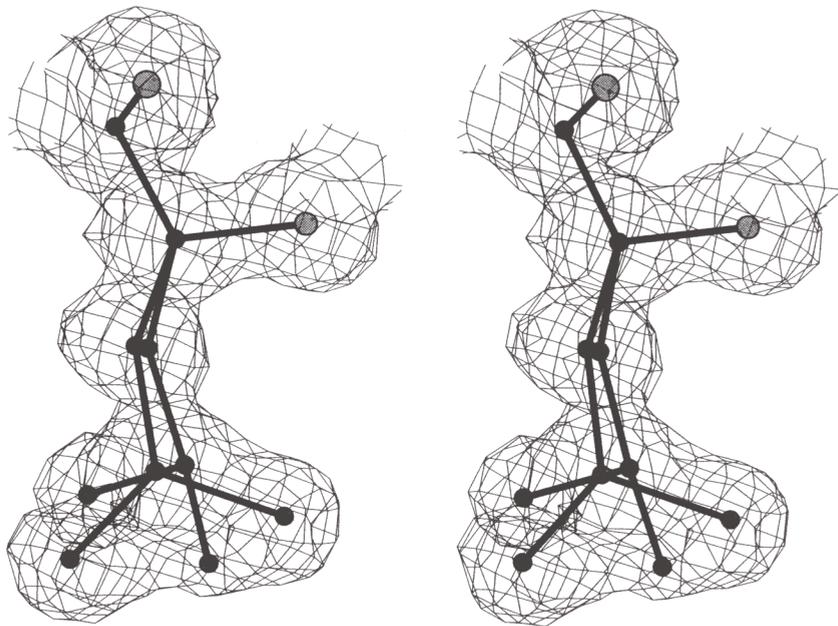


FIG. 3. Various residues with their side chains in double conformations with the $2F_o - F_c$ electron density at 1σ from the structure of *Erwinia* asparaginase (1O7J).

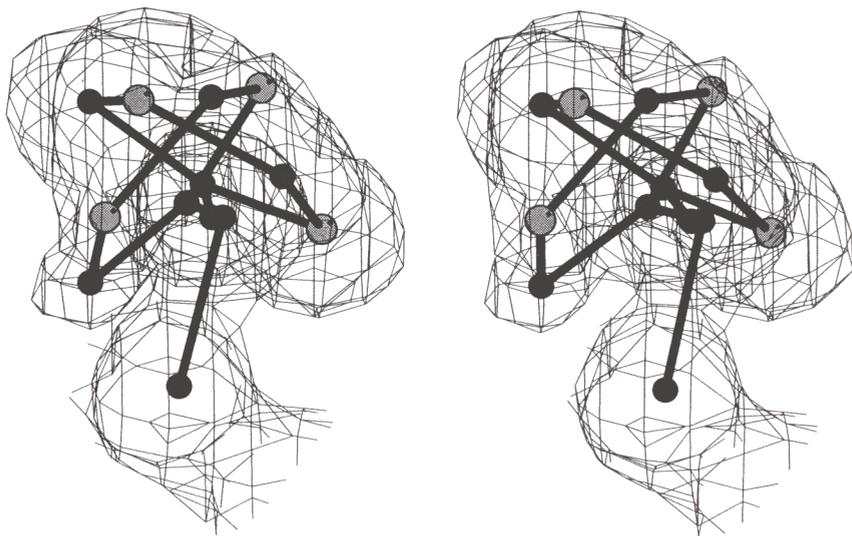


FIG. 4. The His A202 residue of *Erwinia* asparaginase (1O7J) showing two orthogonal conformations of its imidazole ring.

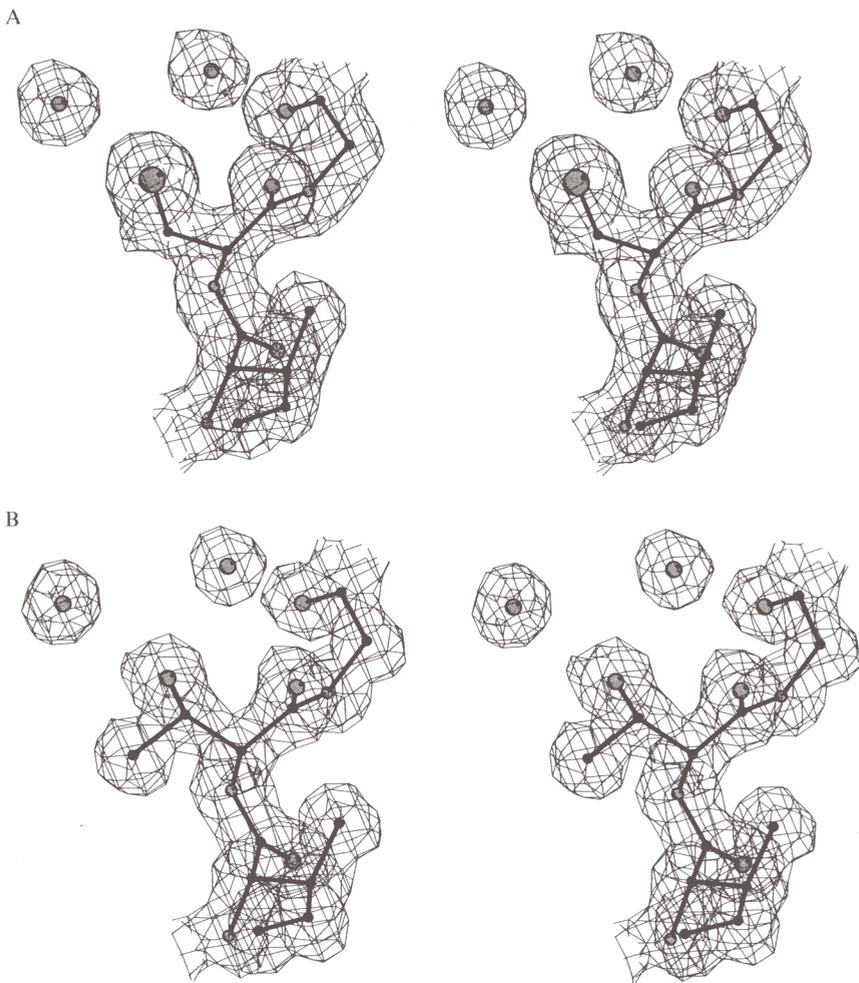


FIG. 5. The residue A72 in RNase Sa with the $2F_o - F_c$ electron density at the 1σ level (A) in the 1.8-Å structure (1SAR),²² modeled as a cysteine, and (B) in the 1.2-Å structure (1RGG),²¹ correctly modeled as a threonine. The clear effect of the model bias in the 1.8-Å structure removed most of the electron density of the second γ atom.

atomic-resolution investigations this opinion must be largely abandoned. Indeed, at “traditional” resolution of about 2 Å or lower, it is not possible to identify hydrogen atoms directly by crystallographic techniques. However, even at medium resolution they contribute to diffraction, and inclusion of hydrogens in the calculated positions (which is possible and advisable) leads to lower R -factors and clearer Fourier maps. Without

hydrogens, their parent atoms will tend to refine toward the center of gravity of the X–H pair. Of course this effect is negligible at low resolution, but becomes significant when data resolution increases. However, experience with small structures suggests that hydrogens cannot be refined as independent atoms even at atomic resolution. They usually are refined as “riding” on their parent atoms with related (e.g., 20 or 50% higher) isotropic B -factors. This procedure is “cheap” as it does not increase the number of refined parameters. Perhaps it is pointless to include hydrogen atoms on residues in multiple, partially occupied conformations.

The interpretability of difference density peaks depends not only on resolution, but also on many other factors, such as the quality of diffraction data, B -factor overall, and in the particular protein region, average level of noise in the Fourier synthesis, etc. As usual with the interpretation of the electron density maps, some peaks are clear whereas others may be very “subjective.”

Among the H-atom difference map peaks, the most clearly visible ones correspond to main chain hydrogens, that is to C $^{\alpha}$ –H and peptide N–H atoms. Usually at atomic resolution a majority of main chain hydrogen atoms have a reasonable difference density, sometimes even more than 75% of them. Hydrogen peaks around side chain atoms are usually less pronounced, but in well-defined protein regions they can be clearly identifiable, even within methyl groups (Fig. 6).¹⁹ Clear difference peaks were identified for all hydrogens in one of the tyrosines (Tyr-13) of rubredoxin in the first atomic resolution least-squares refinement of any protein.²³

Hydrogens within certain hydrogen bonds between two similarly electronegative atoms may sometimes adopt two “tautomeric” positions on either of the two energy minima close to one or another donor/acceptor. Such a split of hydrogen positions has been identified in X-ray structures of smaller molecules (e.g., cyclodextrin), but there is no chance to identify such partially occupied hydrogen sites in protein crystal structures. This consideration is particularly relevant to certain solvent waters, although low temperatures may stabilize one of the alternative directionalities of H-bonding networks.

In a number of particularly well-refined structures water hydrogen atoms have been identified. In the 0.78-Å structure of subtilisin²⁴ 14% of all protein hydrogen atoms have difference density peaks above 3σ , 65% are above the 2σ level, and several water molecules have clear density

²³ K. D. Watenpaugh, T. N. Margulis, L. C. Sieker, and L. H. Jensen, *J. Mol. Biol.* **122**, 175 (1978). K. D. Watenpaugh, L. C. Sieker, and L. H. Jensen, *J. Mol. Biol.* **131**, 509 (1979). K. D. Watenpaugh, L. C. Sieker, and L. H. Jensen, *J. Mol. Biol.* **138**, 615 (1980).

²⁴ P. Kuhn, M. Knapp, M. Soltis, G. Ganshaw, M. Thoene, and R. Bott, *Biochemistry* **37**, 13446 (1998).

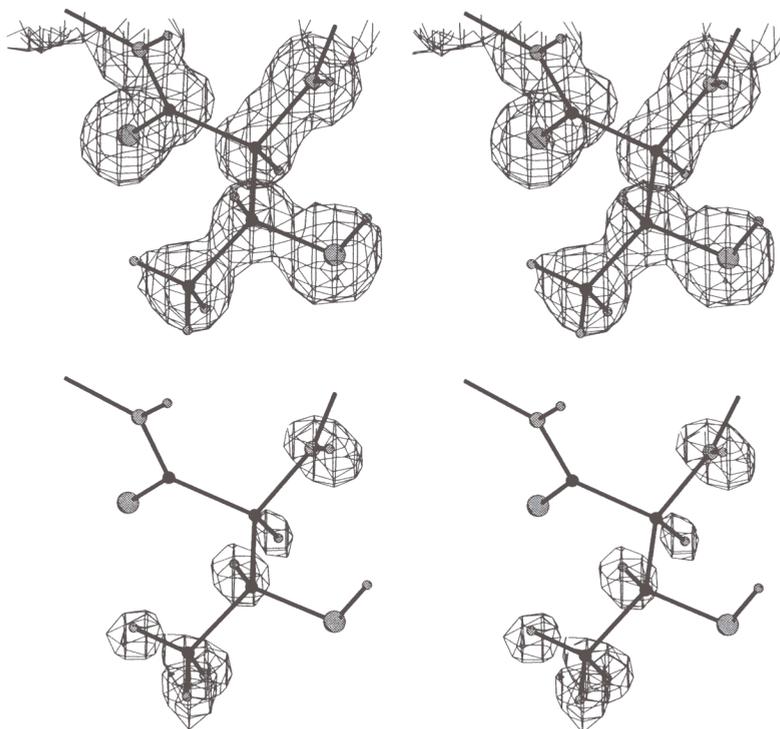


FIG. 6. Residue Thr-71 of ferredoxin (2FDN)¹⁹: (top) the $2F_o - F_c$ electron density at the 1.5σ level and (bottom) the difference map calculated before introduction of hydrogen atoms.

visible for their hydrogen atoms. In the 0.94-Å resolution structure of con-canavalin A²⁵ several waters have both their hydrogens identifiable in the map and even more have only one hydrogen. In the latter case the water molecules are surrounded by several possible H-bond acceptors, and the second hydrogen is most probably associated with more than one acceptor.

Nonclassic H Bonds

The occurrence of the main chain C-H...O hydrogen bonds in proteins and their role in maintaining the integrity of the protein secondary structure have been identified earlier,^{26,27} mainly between C^α-H donors and

²⁵ A. Deacon, T. Gleichmann, A. J. Kalb, H. Price, J. Raftery, G. Bradbrook, J. Yariv, and J. R. Helliwell, *J. Chem. Soc. Faraday Trans.* **1997**, 4305 (1997).

²⁶ Z. S. Derewenda, L. Lee, and U. Derewenda, *J. Mol. Biol.* **252**, 248 (1995).

²⁷ G. F. Fabiola, S. Krishnaswamy, V. Nagarajan, and V. Pattabhi, *Acta Crystallogr. D* **53**, 316 (1997).

carbonyl O atoms within both parallel and antiparallel β -sheet regions. This analysis was based on the statistics of interatomic contacts, taking into account van der Waals radii of the involved atoms, including calculated positions of hydrogen atoms. The atomic-resolution investigations allow one to identify such interactions with enhanced confidence, not only because of more accurately refined atomic positions, but also because of the appearance of hydrogen-atom peaks in the difference Fourier maps.^{28,29} In the structure of haloalkane dehalogenase³⁰ 141 such C–H...O interactions have been found. Interestingly, it has been observed³¹ that in the β -sheet structures the C $^{\alpha}$ –H peaks are located not exactly at the “tetrahedral” positions with respect to their parent C $^{\alpha}$ atoms, but are slightly shifted, so that they can form more effective hydrogen bonds with carbonyl oxygen atoms of the other strand.

However, the hydrogen atoms located on various other carbon atoms may in certain cases form hydrogen bonds with electronegative acceptors. In the structures of ASV integrase³² and proteinase K³³ one of the His residues, apparently protonated in integrase and uncharged in proteinase K, forms hydrogen bonds through all four of its available imidazole ring atoms (N $^{\delta 1}$, C $^{\delta 2}$, C $^{\epsilon 1}$, and N $^{\epsilon 2}$), and there is no indication of any possible disorder involving the 180° flip of the ring, (Fig. 7).

The term “electronegative acceptor” should also be expanded to include π -electron systems of the aromatic rings.³⁴ The early investigation of BPTI identified the interaction between the amide N–H and the tyrosine ring;³⁵ this hydrogen atom was effectively shielded from exchange by deuterium. In fact this tyrosine accepts two N–H... π hydrogen bonds,³¹ from one side the main chain N–H, and from another side the asparagine side chain amide. Both hydrogen atoms are visible in the difference map. In the structure of concanavalin A²⁵ a water molecule placed at a

²⁸ T. Sandalova, G. Schneider, H. Kack, and Y. Lindqvist, *Acta Crystallogr. D* **55**, 610 (1999).

²⁹ L. Esposito, L. Vitagliano, F. Sica, G. Sorrentino, A. Zagari, and L. Mazzarella, *J. Mol. Biol.* **297**, 713 (2000). L. Esposito, G. Vitagliano, A. Zagari, and L. Mazzarella, *Protein Eng.* **13**, 825 (2000).

³⁰ I. S. Ridder, H. J. Rozeboom, and B. W. Dijkstra, *Acta Crystallogr. D* **55**, 1273 (1999).

³¹ A. Addlagatta, S. Krzywdka, H. Czapinska, J. Otlewski, and M. Jaskólski, *Acta Crystallogr. D* **57**, 649 (2001).

³² J. Lubkowski, Z. Dauter, F. Yang, J. Alexandratos, G. Merkel, A. M. Skalka, and A. Wlodawer, *Biochemistry* **38**, 13512 (1999).

³³ C. Betzel, S. Gourinath, P. Kumar, P. Kaur, M. Perbandt, S. Eschenburg, and T. P. Singh, *Biochemistry* **40**, 3080 (2001).

³⁴ G. R. Desiraju and T. Steiner, “The Weak Hydrogen Bond.” Oxford University Press, Oxford, 1999.

³⁵ A. Wlodawer, J. Walter, R. Huber, and L. Sjölin, *J. Mol. Biol.* **180**, 301 (1984).

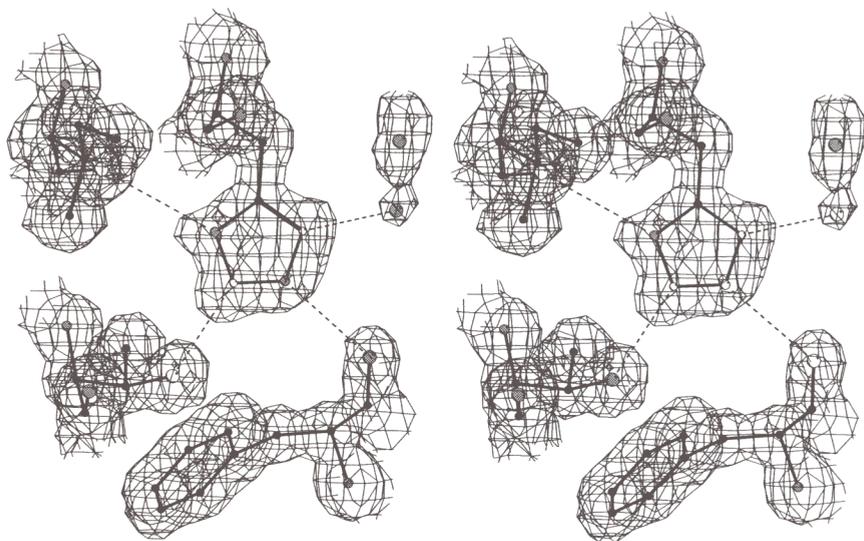


FIG. 7. The surrounding of His-142 in integrase (1CXQ)²⁰ and the $2F_o - F_c$ map at 1σ showing all its imidazole ring atoms engaged in hydrogen bonds.

perpendicular distance of about 3.5 Å from the phenyl ring of a tyrosine donates a hydrogen atom that is clearly visible in the difference map.

The analysis of atomic resolution structures clearly shows the existence of many weak, nonstandard H-bonding interactions within proteins. Their contribution to the integrity of the protein fold is increasingly appreciated.

Protonation of Charged Groups

The most interesting hydrogen atoms from the biochemical point of view are those that may be located on ionizable charged groups and are governed by acid/base equilibrium. Unfortunately, these hydrogens are most difficult to identify in the difference Fourier maps, since they usually reside at termini of amino acid side chains (Asp, Glu, Arg, Lys, His), which tend to have higher than average B -factors. Moreover, such O-H or N-H bonds are polarized, so that effectively the bonding electrons are shifted toward the electronegative, heavier atoms.

However, in several atomic resolution investigations such acidic hydrogen atoms have been identified. In concanavalin A²⁵ a pair of oxygens from neighboring carboxylates at a distance of 2.58 Å are connected by a hydrogen bond, with the difference density corresponding to the hydrogen located close to the aspartate oxygen. Although concanavalin does not have any enzymatic activity, such features are typical for active sites of

the hydrolytic enzymes. Indeed, hydrogen atoms protonating active site residues are obviously most interesting for discussing enzymatic mechanisms. Such crucial hydrogens have been identified in several structures. For example, within the Asp-His-Ser catalytic triad of subtilisin,²⁴ hydroxynitrile lyase,³⁶ elastase,³⁷ and proteinase K³³ the difference density shows the hydrogen located at the N^{δ1} atom of the histidine pointing toward the aspartate carboxyl oxygen. In proteinase K the density of the hydroxyl hydrogen of catalytic serine is also visible pointing toward the N^{ε2} atom of histidine.

In the complex of a glycosidase with the inhibitor containing a sugar-fused imidazole ring,³⁸ the difference density showed that the imidazole N atom (corresponding to normal oligo-sugar glycosidic oxygen) is protonated by the enzymatic glutamate, which helped to clarify the stereochemistry of the enzymatic reaction.

With the high accuracy of the bond length estimation that is possible in very high-resolution structures, the protonation state of the carboxyl functions can be inferred indirectly from the comparison of the two C–O distances. In the charged –COO[–] form, both oxygen atoms are equivalent and both bonds are expected to be equal, whereas in the protonated, neutral –COOH form, the bond to the protonated oxygen atom should be longer than that to the other oxygen. The inspection of the C–O bond lengths within the clusters of the neighboring carboxyl groups in the structure of the serine-carboxyl proteinase³⁹ clearly shows the directionality of the hydrogen bond networks and identifies which atoms are protonated (Fig. 8). In the thorough analysis of the structures of RNase A at six pH values ranging from 5.2 to 8.8⁴⁰ a clear trend was identified in the endocyclic bond angles of the imidazole group of the catalytic histidine, correlated to the changes in pH and the protonation of the imidazole nitrogen atoms.

Validation and Dictionary Targets

The standard practice in refining macromolecules is to use stereochemical restraints as additional observables, supplementing the X-ray data. The geometric properties of the refined model are compared with the target

³⁶ K. Gruber, M. Gugganig, U. G. Wagner, and C. Kratky, *Biol. Chem.* **380**, 993 (1999).

³⁷ M. Würtele, M. Hahn, K. Hilpert, and W. Höhne, *Acta Crystallogr. D* **56**, 520 (2000).

³⁸ A. Varrot, M. Schüleïn, M. Pipelier, A. Vasella, and G. J. Davies, *J. Am. Chem. Soc.* **121**, 2621 (1999).

³⁹ A. Wlodawer, M. Li, A. Gustchina, Z. Dauter, K. Uchida, H. Oyama, N. E. Goldfarb, B. M. Dunn, and K. Oda, *Biochemistry* **40**, 15602 (2001).

⁴⁰ R. Berisio, V. S. Lamzin, F. Sica, K. S. Wilson, A. Zagari, and L. Mazzarella, *J. Mol. Biol.* **292**, 845 (1999). R. Berisio, F. Sica, V. S. Lamzin, K. S. Wilson, A. Zagari, and L. Mazzarella, *Acta Crystallogr. D* **58**, 441 (2002).

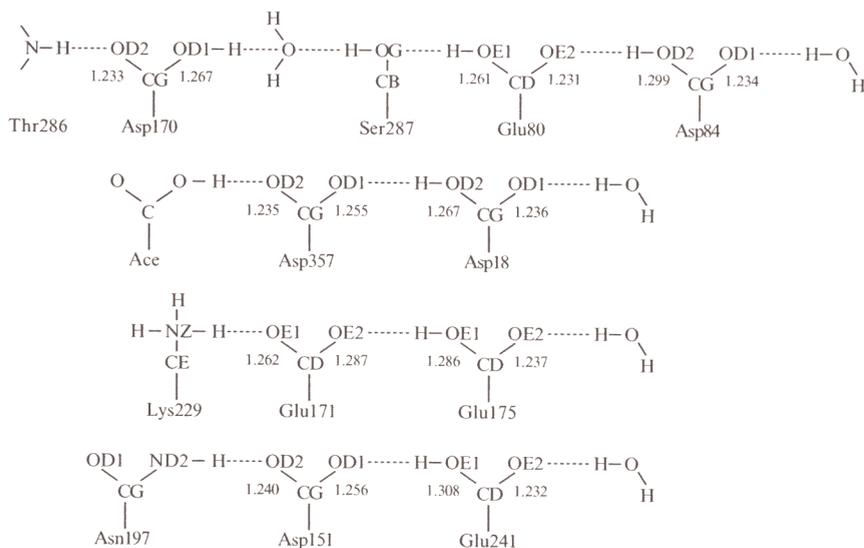


FIG. 8. Four stretches of hydrogen bonded carboxyl functions in PSCP (1GA6).³⁹ The analysis of the C–O bond lengths within these groups defines the directionality of H-bonds within each cluster.

values of various geometric parameters. The most commonly used target values of bond distances and angles were compiled by Engh and Huber⁴¹ from the structures of amino acids and peptides in the Cambridge Structural Database.⁴²

At atomic resolution the number of measured reflections exceeds many times the number of refined atomic parameters. For the crystal containing 50% of solvent and data extending to 1.2 Å resolution there are about five reflections per parameter, including an anisotropic displacement model. For a tightly packed crystal with only 35% of solvent and data to 1.0 Å there are about eight parameters. This ratio of observables per parameter is very similar in small structure crystallography, at about 0.8 Å. Somewhat lower resolution of data from protein crystals is compensated by the presence of the solvent region, which contains no individually refined atoms. In principle therefore, refinement of protein models should not require any restraints. Indeed, well-defined parts of the protein chain with

⁴¹ R. Engh and R. Huber, *Acta Crystallogr. A* **47**, 392 (1991).

⁴² F. H. Allen and V. J. Hoy, "International Tables of Crystallography," Vol. F, p. 663. Kluwer Academic Publishers, Dordrecht, 2001.

low B -factors refine in a stable manner. Only more flexible parts, with high B -factors or partially disordered, require restraints to preserve a reasonable geometry.⁴³

At atomic resolution the contribution of X-ray terms to the refinement procedure is much greater than the influence of restraints. In many instances workers have observed that individual stereochemical parameters refined consistently to values different than in the restraint library! This may identify either unusual, strained conformations, or a deficiency of the target library.

In general, however, refinement of many atomic-resolution structures has confirmed the validity of the Engh and Huber library of bonds and angles; only a few corrections have been postulated. For example, a clear difference between the $N-C^\alpha-C$ angles for residues in α -helices (average $111.8^\circ \pm 1.7^\circ$) and those in β -sheet ($109.1^\circ \pm 1.8^\circ$) was identified in the 0.87-Å structure of RNase A.²⁹ Small revisions were proposed for some bond lengths on the basis of the structure of cutinase⁴⁴ and four other atomic resolution structures.⁴⁵

An important message, from virtually all atomic-resolution investigations, relates to the planarity of peptide groups. Whereas other planar moieties, in particular aromatic rings, preserve strict flatness, peptides often display significant distortion from planarity. Deviations of the $C^\alpha-C-N-C^\alpha$ torsion ω -angle by up to 20° have been observed in many structures.⁴⁶ Such distortions occur as a consequence of the strain in the protein main chain, in tight turns, or when a side chain is anchored by strong interactions with neighboring groups. Overrestraining peptide planarity may prevent neighboring atoms from adopting their correct conformation. It is recommended that the weight of the peptide-planarity restraint should produce a standard deviation from the perfectly planar value of about 6° , instead of the often targeted 3° .

The ω -angle is not the only measure of peptide planarity, since the carbonyl O atom should also in an ideal case lie in the same plane. The degree of its distortion from the plane can be measured by the pyramidalization of the carbonyl C atom, expressed as the difference of the appropriate torsion angles $\tau(C^\alpha-C-N-C^\alpha) - \tau(O-C-N-C^\alpha) - 180^\circ$. The analysis of

⁴³ Z. Dauter, L. C. Sieker, and K. S. Wilson, *Acta Crystallogr. B* **48**, 42 (1992).

⁴⁴ S. Longhi, M. Czjzek, V. Lamzin, A. Nicolas, and C. Cambillau, *J. Mol. Biol.* **268**, 779 (1997).

⁴⁵ V. S. Lamzin, Z. Dauter, and K. S. Wilson, *J. Appl. Crystallogr.* **28**, 338 (1995).

⁴⁶ K. S. Wilson, S. Butterworth, Z. Dauter, V. S. Lamzin, M. Walsh, S. Wodak, S. Pontius, J. Richelle, A. Vaguine, C. Sander, R. W. W. Hooft, G. Vriend, J. M. Thornton, R. A. Laskowski, M. W. MacArthur, E. J. Dodson, G. Murshudov, T. J. Oldfield, T. Kaptein, and J. A. C. Rullman, *J. Mol. Biol.* **276**, 417 (1998).

peptide bonds in eight atomic-resolution structures shows clear negative correlation between the peptide C–N and C=O bond lengths.²⁹

The geometric parameters, which were restrained during refinement to the library target values, have little value as validation tools,^{11,47} since they often reproduce the assumed mean and dispersion of the restrained parameters. A common example is the deviation of bond lengths from target values of 0.010 Å or less, which only proves that the model geometry is unnecessarily overrestrained. Atomic resolution structures suggest that the desired value of this deviation should be about 0.020 Å to reflect the true, natural variation!

The best parameters to use for validation are those that were not restrained during model refinement. Most useful is the Ramachandran plot,⁴⁸ since the main chain conformational torsion angles ϕ and ψ are not restrained; similarly the side-chain rotamer angles χ can be used. In atomic resolution structures the ϕ , ψ angles are tightly clustered in the core α and β regions of the Ramachandran plot, and even suggest a possible redefinition of these regions.⁴⁶

In the context of structure validation, it is worth mentioning the accuracy of the estimation of crystal cell parameters. At atomic resolution inaccuracy of cell dimensions may have a significant adverse effect on the accuracy of the protein model.⁴⁶ This is not a trivial problem, since the cell dimensions estimated from diffraction-data processing are correlated with the knowledge of the X-ray wavelength and possibly of the crystal-to-detector distance. If the experimenter is very careful, the wavelength can be estimated with as good as 0.01% accuracy ($\lambda = 1.5418$ Å for $\text{CuK}\alpha$, is this accurate), but it is difficult, though possible, to measure wavelength as accurately as this at a synchrotron source. The experimenter should beware! Some, but not all data-reduction programs are able to refine the unit cell parameters in a way that is independent of the precise measure of the specimen-to-detector distance. Therefore the accuracy of cell parameters could be as good as about 0.02% or worse than 1%. Moreover, particularly for lower symmetry crystals, the cell dimensions refined from reflection positions vary depending on the crystal orientation. The postrefinement procedure may average them, but only with limited accuracy. Cell parameters should be quoted with true precision, and examples such as $a = 123.456$ Å or $\beta = 123.45^\circ$ are not realistic.

⁴⁷ G. J. Kleywegt, *Acta Crystallogr. D* **56**, 249 (2000).

⁴⁸ G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *J. Mol. Biol.* **7**, 95 (1963).

Valence Electrons

In the standard crystallographic practice, the scattering of individual atoms is expressed by spherical scattering functions. This gives a good agreement with experiment, even in small-structure crystallography where the resolution of diffraction data reaches about 0.8 Å. However, whereas the inner core electrons of the closed shells are spherically symmetric, the valence electrons of the outer shells, which are involved in interatomic bond formation, do not conform with the spherical approximation. This effect is very subtle, and to visualize the deformation of electron density by the X-ray diffraction experiment, and to decouple it from the anisotropic atomic displacement effect, one needs particularly accurate intensity measurements at very high resolution and at low temperature.⁴⁹ Even with small-structure crystals, such charge-density deformation studies are not routine.

The constituent atoms and bonds of proteins have the same properties as in any other molecule, including the symmetry of their valence electron shells. Protein crystals do not diffract well enough to substantiate such analysis of individual atomic electron clouds. However, they contain multiple copies of the same building block, that is a (trans) peptide group, consisting of five planar atoms between a pair of C α s of consecutive residues along the protein chain. The averaging of electron density of N multiple copies of well-defined peptides should diminish the map noise square root (N) times. This procedure applied to crambin and subtilisin⁵⁰ refined with a classic, spherical atomic model showed clear deformation densities revealing σ -bonding electrons. More thorough calculations based on crambin data extending to 0.54 Å,⁵¹ which involved representing the atomic scattering parameters by the multipolar pseudoatom model, led to a satisfactory modeling of the electron-density distribution. Such experiments can provide unbiased information on the distribution of charge in proteins, which is crucial, e.g., for studies of substrate binding or receptor recognition.

Metal Coordination

Heavier atoms, such as metals, contribute more than light atoms to the total crystal scattering, therefore their parameters are refined in a least-squares procedure with higher precision. C, N, and O atoms located

⁴⁹ P. Coppens, "X-Ray Charge Densities and Chemical Bonding." Oxford University Press, Oxford, 1997.

⁵⁰ V. S. Lamzin, R. J. Morris, Z. Dauter, K. S. Wilson, and M. M. Teeter, *J. Biol. Chem.* **274**, 20753 (1999).

⁵¹ C. Jelsch, M. M. Teeter, V. Lamzin, V. Pichon-Lesme, B. Blessing, and C. Lecomte, *Proc. Natl. Acad. Sci. USA* **97**, 3171 (2000).

in rigid parts of the protein chain may have the estimated standard uncertainties of their positions in the order of 0.01–0.02 Å. The positional accuracy of metals may be an order of magnitude better, comparable with the accuracy of small structures. Interatomic distances within the $[\text{Fe}_4\text{-S}_4]$ clusters in ferredoxin refined at a resolution higher than 1.0 Å^{19,52} are estimated with SUs in the range 0.002–0.004 Å, clearly showing that the clusters are significantly distorted from the ideal tetrahedral geometry, but each in a different way. These distortions, exerted by the surrounding protein, are important for discussing the redox and electron transfer properties of ferredoxin.

In the 0.94-Å resolution structure of concanavalin A,²⁵ the accuracy of bonds between Mn^{2+} and Ca^{2+} ions and coordinating atoms is 0.005–0.010 Å, and that of the corresponding angles is about 0.2°. The observed distortions from the ideal octahedral coordination of Mn^{2+} explain the EPR and CD spectra of this transition metal binding site. The Ca^{2+} ion has seven oxygen ligands, asymmetrically placed at distances 2.360–2.513 Å where the carboxylate oxygen atom most distant from Ca^{2+} is shared with Mn^{2+} .

The high-precision model of ZnS_4 and FeS_4 coordination was obtained by refining two structures of rubredoxin.⁵³ Similar sites are present in various other proteins, e.g., zinc-fingers. The metal coordination by four cysteine sulfurs corresponds to a tetrahedron flattened along the 2-fold axis, which also relates the surrounding parts of the protein chain, including the characteristic CXXC motif.

A zinc ion is present in the deuterolysin,⁵⁴ coordinated by two histidines, an aspartate, and two water molecules in a distorted octahedral geometry. Alcohol dehydrogenase contains two zinc atoms. The recent atomic-resolution analysis⁵⁵ showed some unexpected features around the catalytic Zn site, and led to a proposed revision of the enzymatic mechanism, involving the puckering of NADH coenzyme and involvement of the fifth Zn ligand, postulated as a hydroxide ion.

Since the solution of first protein X-ray structures of myoglobin and hemoglobin in the 1950s much attention was directed toward the heme iron coordination geometry and electronic states. Several atomic resolution structures of heme-containing proteins are now available.

⁵² K. Fukuyama, T. Okada, Y. Kakuta, and Y. Takahashi, *J. Mol. Biol.* **315**, 1155 (2002).

⁵³ Z. Dauter, K. S. Wilson, L. C. Sieker, J. M. Moulis, and J. Meyer, *Proc. Natl. Acad. Sci. USA* **93**, 8836 (1996).

⁵⁴ K. E. McAuley, Y. Jia-Xing, E. J. Dodson, J. Lehmbeck, P. R. Oestergaard, and K. S. Wilson, *Acta Crystallogr. D* **57**, 1573 (2001).

⁵⁵ R. Meijers, R. J. Morris, H. W. Adolph, A. Merli, V. S. Lamzin, and E. S. Cedergren-Zeppezauer, *J. Biol. Chem.* **276**, 9316 (2001).

The analysis of the sperm whale myoglobin structures in aquomet, deoxy, and carbonmonoxy states⁵⁶ showed that CO is coordinated to iron in nearly linear conformation, bent by about 7°, in contrast to O₂, which binds in a more angular fashion. Binding of carbon monoxide is weakened by a steric clash between its oxygen and a histidine and a valine from helix E, and a need for reorganization of the heme, iron atom, and helices E and F.

In the other thorough atomic resolution work on sperm whale myoglobin in four states, carbonmonoxy, aquomet, oxy, and deoxy,⁵⁷ it was found that CO is coordinated at a small angle of 9°, whereas O₂ is bent by 58°. The distal histidine ring in MbCO and MbO₂ was found in two similar conformations, but in MbCO a third, swung-out conformation with about 20% occupancy was also observed. The analysis of anisotropic displacement parameters suggests that the heme and the proximal helix F move as rigid bodies, in contrast to the more flexible distal helix E. Details of the conformational substates are discussed and correlated with the spectroscopic and other available data.

The trematode hemoglobin⁵⁸ has a tyrosine instead of a typical distal histidine at E7 position. Its exceptionally high oxygen affinity is explained by the stabilization of the ligand by a hydrogen bonding interaction with another tyrosine at position B10.

The structures of nitrophorin 4 complexed with NO and NH₃ showed that the iron is in the ferric Fe(III) state with ruffled heme and bent Fe–N–O geometry.⁵⁹ Atomic resolution accuracy allowed a detailed discussion of the relations between the iron electronic state and redox properties and small distortions of the heme and other Fe ligands.

Several structures of various heme-containing cytochromes have been refined at atomic resolution. The structure of cytochrome c₆,⁶⁰ solved *ab initio*, provided the most accurate model of a heme protein at a time. Its structure has been compared with plastocyanin, which is the alternatively used protein in the electron transfer from photosystem I in primitive organisms. The single electron transfer pathway, serving interactions with both

⁵⁶ G. S. Kachalova, A. N. Popov, and H. D. Bartunik, *Science* **284**, 473 (1999).

⁵⁷ J. Vojtechovsky, K. Chu, J. Berendzen, R. M. Sweet, and I. Schlichting, *Biophys. J.* **77**, 2153 (1999).

⁵⁸ A. Pesce, S. Dewilde, L. Kiger, M. Milani, P. Ascenzi, M. C. Marden, M. L. Van Hauwaert, J. Vanfleteren, L. Moens, and M. Bolognesi, *J. Mol. Biol.* **309**, 1153 (2001).

⁵⁹ S. A. Roberts, A. Weichsel, Y. Qiu, J. A. Shelnutt, F. A. Walker, and W. R. Montfort, *Biochemistry* **40**, 11327 (2001).

⁶⁰ C. Frazão, C. M. Soares, M. A. Carrondo, E. Pohl, Z. Dauter, K. S. Wilson, M. Hervás, J. A. Navarro, M. A. De la Rosa, and G. M. Sheldrick, *Structure* **3**, 1159 (1995).

cytochrome *f* and photosystem I, is proposed for cytochrome *c*₆, in contrast to two separate pathways in plastocyanin.

The structure of a soluble cytochrome *c*₅₅₃ was solved *ab initio* using ARP and refined at 0.97 Å resolution.⁶¹ Its very low reduction potential of +47 mV was discussed in relation to the high solvent accessibility of heme.

Methyl-coenzyme M reductase contains the cofactor F₄₃₀, a nickel-containing porphinoïd moiety. The structure of this methane-generating enzyme was refined at 1.16 Å⁶² in complex with coenzymes M and B. Interactions of the protein and its several ligands and bound ions were discussed in detail and the catalytic mechanism proposed.

Waters and Other Solvent Molecules

The satisfactory modeling of the solvent–water molecules is difficult at medium resolution. This problem persists also at atomic resolution. In general, crystals able to diffract to atomic resolution tend to be closely packed and have less than an average amount of solvent but, except for small oligopeptides or antibiotics, they also contain bulk solvent regions corresponding to the completely disordered liquid phase. These structures have more well-defined water sites, which are stable during refinement and practically belong to the well-defined protein model, but most troublesome remains the intermediate shell of poorly defined or partially occupied water sites at the border of the bulk solvent region.

The problems with Fourier synthesis in this intermediate region, located at the border between the well-defined protein with good waters, modeled as discrete atomic sites, and bulk solvent, represented as a constant level of electron density, is responsible for the high *R*-factor at the lowest resolution, often observed even for structures refined at atomic resolution. The solvent *B*-factor parameter⁶³ partially takes into account the smooth conversion of these two representations.

Only crystals of relatively small polypeptides do not have bulk solvent, and there it is possible to model all solvent–water molecules. The largest structure where practically all expected waters were modeled is crambin. In the investigation of the nonhomogeneous (Ser/Pro-22-Ile/Leu-25) form of crambin at 0.83 Å resolution⁶⁴ 141 water sites were identified and their

⁶¹ S. Benini, A. Gonzalez, W. R. Rypniewski, K. S. Wilson, J. J. van Beeumen, and S. Ciurli, *Biochemistry* **39**, 13115 (2000).

⁶² W. Grabarse, F. Mahlert, E. C. Duin, M. Goubeaud, S. Shima, R. K. Thauer, V. Lamzin, and U. Ermler, *J. Mol. Biol.* **309**, 315 (2001).

⁶³ D. E. Tronrud, *Methods Enzymol.* **277**, 306 (1997).

⁶⁴ M. M. Teeter, S. M. Roe, and N. H. Heo, *J. Mol. Biol.* **230**, 292 (1993).

occupancies and B -factors refined, with the total amount equivalent to 97% of the 91 waters expected from the crystal density, and in addition one molecule of ethanol was found. The pure sequence (Ser-22/Ile-25) structure refined at 150 K and 0.89 \AA^{65} revealed correlated alternative networks of disordered side chains and complementary water sites. The comparison of crambin structures at eight temperatures between 100 and 293 K⁶⁶ identified the transition of the hydrated protein into the glassy state at about 180 K, in which multiple conformations of protein side chains and five- or six-membered rings of water sites are dynamically coupled.

Usually all solvent sites are modeled as waters, in spite of the fact that often the mother liquor contains various other ionic or organic additives. Ions, such as NH_4^+ , Na^+ , and Mg^{2+} are isoelectronic with H_2O and it is practically impossible to identify them in the electron density. The only way to verify the identity of Mg^{2+} and Na^+ is through their characteristic coordination, Mg^{2+} rather strictly octahedral and Na^+ irregular pentahedral. However, if such sites are placed outside of the first, well-defined solvent shell, but are partially disordered together with the coordinating waters, there is no chance to characterize them even in atomic resolution structures. A very careful analysis of charges of the hydrogen bonding groups in the 1- \AA structure of pheromone Er-1⁶⁷ led to the classification of three solvent sites as NH_4^+ ions.

Similar difficulties arise with such organic molecules as glycol, polyethylene glycol (PEG), methylpentanediol (MPD), glycerol, or molecules of buffers, such as Tris and HEPES. All these compounds are polar, well soluble in water, and able to form hydrogen bonds. It is possible to identify and refine such molecules only if they are anchored by hydrogen bonds at the protein surface or sometimes between two protein molecules in the crystal lattice, otherwise they tend to be completely disordered within the bulk solvent. In several atomic resolution structures such molecules have been identified, more often than at medium resolution. Two glycerol molecules, which diffused into the crystal during a short soak in cryoprotecting buffer, were refined in the active site of dUTPase.⁶⁸ Similarly, two cryoprotectant ethylene glycol molecules were present in the crystal of the BPTI mutant.³¹ One or two molecules of MPD were found in the biotin binding

⁶⁵ A. Yamano, N. H. Heo, and M. M. Teeter, *J. Biol. Chem.* **272**, 9597 (1997).

⁶⁶ M. M. Teeter, A. Yamano, B. Stec, and U. Mohanty, *Proc. Natl. Acad. Sci. USA* **98**, 11242 (2001).

⁶⁷ D. H. Anderson, M. S. Weiss, and D. Eisenberg, *Acta Crystallogr. D* **52**, 469 (1996). D. H. Anderson, M. S. Weiss, and D. Eisenberg, *J. Mol. Biol.* **273**, 479 (1996).

⁶⁸ A. Gonzalez, E. Cedergren, G. Larsson, and R. Persson, *Acta Crystallogr. D* **57**, 767 (2001).

sites in each of the four independent protein molecules in the 1.14-Å structure of streptavidine mutant,⁶⁹ and this feature was not seen in lower resolution (1.8–2.0 Å) structures of streptavidin crystals, always grown from 50% MPD solution. The atomic resolution structure of scorpion toxin⁷⁰ permitted identification of both enantiomeric forms of MPD. Both forms of MPD were also identified in the centrosymmetric crystal of α -helical peptide at 0.75 Å resolution.⁷¹ In the structure of ASV integrase³² well-defined molecules of HEPES and citrate are present, both originating from the buffer used in protein crystallization.

B-Factors

At atomic resolution not only positional parameters, but also atomic displacement parameters (ADPs), represented by *B*-factors, can be refined with high accuracy (note: “A” in ADP means atomic, not anisotropic, so that there are isotropic or anisotropic ADPs⁷²). Beyond the 1.2 Å limit, the number of measured intensities justifies the anisotropic modeling of atomic displacements represented by a symmetric tensor with six unique terms. Almost all atomic-resolution crystal structures are refined anisotropically, since this procedure leads to more accurate models and better agreement with experiment, as documented by a significant drop in the R_{free} value. After refinement of anisotropic ADPs the Fourier maps, particularly the difference syntheses, appear much cleaner and have less noise. A thorough discussion of anisotropic ADPs in proteins was published recently.⁷³

Similarly to the situation with positional parameters, the well-defined atoms usually do not require their anisotropic *B*-factors to be restrained, but ADPs of atoms in more flexible parts or in disordered regions, and in particular weak water sites, do not refine well without restraints. In contrast to the positional or geometric parameters, *B*-factors cannot be referred to a predefined library. Various types of anisotropic ADP restraints are implemented in SHELXL: DELU rigid bond restrains the ADP components along the bond between two atoms to be equal, SIMU makes all ADP components of two close atoms similar, and ISOR does not allow ADPs to become too anisotropic.

⁶⁹ S. Freitag, I. Le Trong, L. A. Klumb, P. S. Stayton, and R. E. Stenkamp, *Acta Crystallogr. D* **55**, 1118 (1999).

⁷⁰ B. Zhao, M. Carson, S. E. Ealick, and C. E. Bugg, *J. Mol. Biol.* **227**, 239 (1992).

⁷¹ W. R. Patterson, D. H. Anderson, W. F. Degrado, D. Cascio, and D. Eisenberg, *Protein Sci.* **8**, 1410 (1999).

⁷² D. Schwarzenbach, S. C. Abrahams, H. D. Flack, E. Prince, and A. J. C. Wilson, *Acta Crystallogr. A* **51**, 565 (1995).

⁷³ E. A. Merritt, *Acta Crystallogr. D* **55**, 1109 (1999).

In several atomic-resolution structures the displacement parameters were analyzed particularly thoroughly. It has been found that in most atomic-resolution structures the average anisotropy (the ratio of the minimum to maximum eigenvalue of the U_{ij} tensor, E_{11}/E_{33}) is about 0.40–0.50,⁷³ which shows a significant difference from an isotropic displacement model.

The correlation of B -factor with the square of the distance from the molecular center was found for two lysozymes⁷⁴ and for α -lactalbumin⁷⁵ and, earlier, for a trypsin-like proteinase.⁷⁶ This suggests that to a high degree the molecules follow the rigid body dynamics. The TLS analyses confirmed the validity of such a model for molecules or individual domains.^{77,78}

Structure Solution by Direct or *Ab Initio* Methods

Several atomic-resolution structures have been solved by direct methods or the *ab initio* approach. These two terms are often incorrectly used synonymously. According to the current consensus,⁷⁹ direct methods employ probabilistic relationships among structure factors for solving the phase problem, whereas *ab initio* methods use only native data for this purpose, but not the isomorphous or anomalous differences, for example. Of course, the structure may be also solved *ab initio* by direct methods. Direct methods are often used to find the substructure of heavy (anomalously scattering) atoms, which can be expanded, e.g., by probabilistic phase estimation of Fourier recycling.

The programs most often used for direct methods applications to macromolecules are SnB⁸⁰ and SHELXD.⁸¹ Both are based on the Shake-and-Bake dual-space recycling,⁸² iteratively refining phases in reciprocal space, and selecting atoms in real space. Among 29 atomic resolution structures solved directly or *ab initio*, SnB has been used 10 times and

⁷⁴ K. Harata, Y. Abe, and M. Muraki, *Proteins* **30**, 232 (1998).

⁷⁵ K. Harata, Y. Abe, and M. Muraki, *J. Mol. Biol.* **287**, 347 (1999).

⁷⁶ T. R. Schneider, in "Macromolecular Refinement," Proceedings of the CCP4 Study Weekend (E. Dodson, M. Moore, A. Ralph, and S. Bailey, eds.), p. 133. CLRC Daresbury Laboratory, 1996.

⁷⁷ M. Vlassi, Z. Dauter, K. S. Wilson, and M. Kokkinidis, *Acta Crystallogr. D* **54**, 1245 (1998).

⁷⁸ M. A. Wilson and A. T. Brünger, *J. Mol. Biol.* **301**, 1237 (2000).

⁷⁹ I. Uson and G. M. Sheldrick, *Curr. Opin. Struct. Biol.* **9**, 643 (1999).

⁸⁰ R. Miller, S. M. Gallo, H. G. Khalak, and C. M. Weeks, *J. Appl. Crystallogr. B* **27**, 613 (1994).

⁸¹ T. R. Schneider and G. M. Sheldrick, *Acta Crystallogr. D* **58**, 1772 (2002).

⁸² R. Miller, G. T. DeTitta, R. Jones, D. A. Langs, C. M. Weeks, and H. A. Hauptman, *Science* **259**, 1430 (1993).

SHELXD 9 times. Four times SHELXS⁴ has been used and three times ARP⁸³ expanded the whole structure from heavy atom positions. Two times the recently introduced program ACORN⁸⁴ has been used.

It is well known that the presence of heavier atoms enhances the chance of solving the structure by direct methods.⁸⁵ Indeed, out of 28 atomic resolution structures solved using a single wavelength data, 11 contain metals and only three small structures have no atom heavier than oxygen. The size of these structures varies from a small toxin with 13 residues⁸⁶ to alcohol dehydrogenase containing 7120 atoms.⁵⁵

Conclusion

The number of very-high-resolution structures currently deposited in the PDB clearly shows that protein crystals diffracting to atomic resolution are becoming less rare. Although the primary requirement is the ability of a macromolecule to form a highly ordered crystal, the very significant factor in the recent explosion of the atomic resolution analyses is the progress in all involved techniques. The methods of protein purification and crystal growth, availability of bright synchrotron beam lines, and progress in software used in all stages of the analysis make such projects more tractable. The enhanced accuracy of the resulting models and the possibility of describing very fine structural features are very important for better understanding of the chemistry responsible for the biological properties of macromolecules, which is the ultimate goal of structural biology.

⁸³ A. Perrakis, R. J. Morris, and V. S. Lamzin, *Nat. Struct. Biol.* **6**, 458 (1999).

⁸⁴ J. Foadi, M. M. Woolfson, E. J. Dosdon, K. S. Wilson, Y. Jia-Xing, and Z. Chao-de, *Acta Crystallogr. D* **56**, 1137 (2000).

⁸⁵ G. M. Sheldrick, Z. Dauter, K. S. Wilson, H. Hope, and L. C. Sieker, *Acta Crystallogr. D* **49**, 18 (1993).

⁸⁶ T. Sato, H. Ozaki, Y. Hata, Y. Kitagawa, Y. Katsube, and Y. Shimonishi, *Biochemistry* **33**, 8641 (1994).