Protein Structure Determination Using A Combination of Comparative Modeling and NMR Spectroscopy. Application to the Response Regulator Protein, Spo0F

Brent L. Podlogar,* Gregory C. Leo, Patricia A. McDonnell, Deborah A. Loughney, Gary W. Caldwell, and John F. Barrett

The R. W. Johnson Pharmaceutical Research Institute, Route 202, P.O. Box 300, Raritan, New Jersey 08869

Received March 10, 1997[⊗]

A practical combination of comparative modeling and NMR spectroscopy was used to generate a three-dimensional structure of the response regulator protein, Spo0F. The backbone structure obtained compares to the Spo0F Y13S mutant X-ray structure with an rmsd of 2.0 Å. We provide results which suggest that structures obtained by this method are suitable for drug discovery. The results of the GRID and DOCK methods as applied to the model and X-ray structures of Spo0F are remarkably similar and tend to suggest the same design conclusions. This trend is illustrated by these same techniques applied to two experimentally derived structures of the analogous protein, CheY, which exhibit a pairwise $rmsd_{BB}$ on the same order as that found for the two Spo0F structures.

Introduction

A three-dimensional structure of the key protein in a targeted biological pathway is extremely useful in the design of therapeutic agents, and to date, NMR spectroscopy and X-ray crystallography represent the only two biophysical methods that yield high-resolution structures. Experimental complexities that surround these methods extend the length of time in which structures are solved and often exceeds the time allotted in a pharmaceutical research environment. For NMR-derived structures, the secondary structures of proteins are generally accessible from native and or ^{15}N-labeled proteins and can be obtained quickly.[1] In contrast, a high- resolution solution structure ($rmsd_{BB}$ ~0.5 Å)[2] requires generating a large set of unambiguously assigned NOE constraints (>16 NOEs per residue).[3] The experimental difficulties involved with X-ray crystallography (crystallization conditions, diffraction, the phase problem) are well-known.[4]

Spo0F, the response regulator protein (RR) of the KinA/Spo0F two-component phosphotransfer signal transduction pathway[5−7] was targeted as key enzyme in our efforts to design novel antibacterial therapeutic agents. While efforts were underway to determine the structure of Spo0F using NMR spectroscopy and X-ray crystallography, the possibility of delivering a robust structural model earlier in the drug design process prompted the examination of computational methods. Spo0F shares similar function and sequence identity with nearly 100 RRs, and the degree of sequence similarity among these proteins,[8] a class that possesses a well-defined $(\beta\alpha)_5$ folding motif, permits the global fold to be approximated computationally by comparative modeling techniques.[9] In conjunction with secondary structural information obtained from preliminary NMR data, structures consistent with experimental data known to date can be produced in short order and then integrated into the drug discovery process.

We have applied this approach to Spo0F to generate a three-dimensional structural model using a list of NOE distances obtained from NMR spectroscopy.[10] Despite the large number of RRs identified, only CheY, the homologous RR for adaptive chemotaxis, has been characterized by both NMR and X-ray crystallography.[11−16] NTRC, the RR for nitrogen fixation, has been characterized by NMR.[17] CheY will therefore serve as the reference, or template, for the comparative modeling procedure. The accuracy of the Spo0F model was measured against the X-ray structure of the Y13S Spo0F mutant, which was determined independently during the course of this study by Varughese and coworkers.[18]

Results and Discussion

The tertiary fold of Spo0F has been approximated using the commercially available package MODELER.[19] In MODELER, information such as the amino acid residue, main-chain conformation, side-chain conformation, and local similarity (alignment) between two sequences are defined as restraints in a variable target function comprised of probability density functions for the respective restraints.[20] The sequence alignment of Spo0F and CheY (shown below) yields 24% sequence identity and 57% amino acid similarity. Key residues of the catalytic core in CheY are paired with those of Spo0F and are underlined.

```
SPOOF: -MMNEKILIVDDQYGIRILLNEVFNKEGYQTFQAA-NGLQALDIVTKER
CHEY : ADKELKFLVVDDFSTMRRIVRNLLKELGFNNVEEAEDGVDALNKLQAGG

SPOOF: PDLVLLDMKIPGMDGIEILKRMKVIDE--NIRVIIMTAYGELDMIQESK
CHEY : FGFIISDWNMPNMDGLELLKTIRADSAMSALPVLMVTAEAKKENIIAAA

SPOOF: ELGALTHFAKPFDIDEIRDAVKKYLPLKS
CHEY : QAGASGYVVKPFTAATLEEKLNKIFEKLGM
```

MODELER is highly automated, and all default parameters were utilized; five model structures were generated by this procedure. Ramachandran plots, 3D profiles analyses, and rotomer library analyses indicated that all model structures were consistent with valid protein structures.[21] Visual inspection of the five predicted models revealed that they all retained identical tertiary folds, and pairwise rms deviation for *all* atoms ($rmsd_{ALL}$) were found to be less than 1.5 Å. The close similarity among the comparative molecular mod-

[⊗] Abstract published in *Advance ACS Abstracts,* October 1, 1997.

S0022-2623(97)00156-8 CCC: $14.00

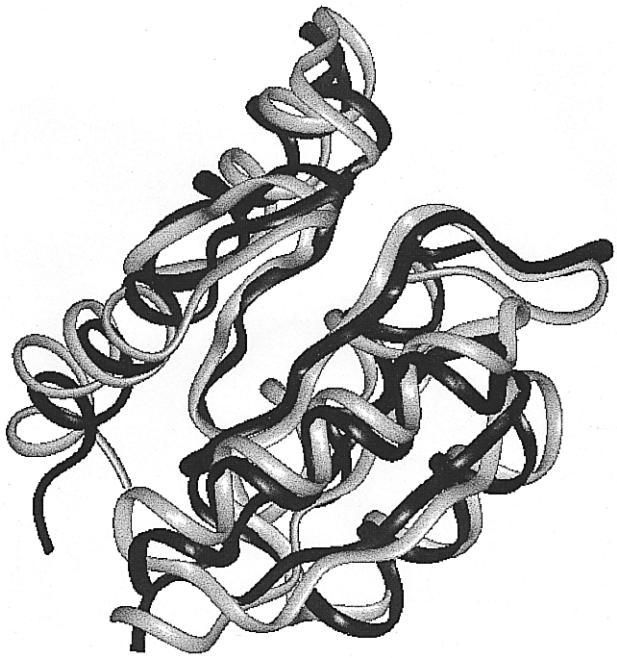

Figure 1. Overlay of the ribbon representations of the comparative molecular model of Spo0F (gray) and the X-ray structure (black).

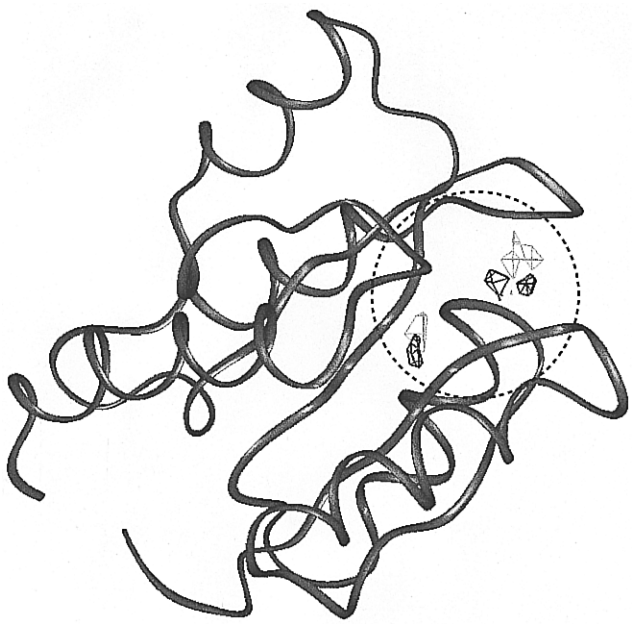

Figure 2. The GRID contour map of CheY for the $N3^+$ probe atom. For clarity, only the ribbon trace of the NMR structure is shown. The bold contour (black) is that resulting from the NMR structure, the thin contour (gray) from the X-ray structure. The circled region denotes the active site.

els prompted the arbitrary use of the first model structure in the refinement procedure. The model compared to the X-ray of the Y13S Spo0F mutant with a pairwise $rmsd_{BB}$ of 2.1 Å.

The NMR-derived distances (332) and dihederal constraints (181) were taken from a recent report[7] and compared against the model structure. Of these, 51 distance restraints were found to be violated (i.e. >0.5 Å). The X-ray structure of the Y13S Spo0F mutant itself violated 24 of the NMR constraints. The model was refined using the distance restraints with a restrained molecular mechanics minimization employing 2.5 kcal/(mol Å^2) restraining force constants.[22] The structure obtained from the refinement compared only slightly better to the X-ray structure of the Y13S Spo0F mutant ($rmsd_{BB}$ = 2.0 Å), but all distance constraints were satisfied. The refined model was considered a valid structure in the same manner as described above and is compared to the X-ray structure in Figure 1. The comparative modeling protocol defines the tertiary fold; the NMR data is able to refine the β-sheet and helical regions; and the force field defines the side-chain conformations. The β-sheets flank helices 1−4, and the network of interactions confer structure upon these helices.

The crystal structure of Y13S Spo0F mutant is one of three isoforms arbitrarily taken from the unit cell and itself possesses an $rmsd_{BB}$ of 0.7 and 0.4 Å compared to the other isoforms, respectively. Variation in experimentally derived structures is common.[4] For example, the crystal structure of CheY has been solved to 1.8 Å resolution,[11] and the solution structure (46 structures) has been solved using 1545 distance constraints from heteronuclear NMR.[15] The X-ray structure and 46 NMR structures are all useful for design efforts, even though they differ among each other with an $<rmsd_{BB}>$ of 1.65 Å.

We have applied the GRID[23] and DOCK[24] methods to the refined Spo0F model and Y13S Spo0F mutant and have found that the differences in their structures have an imperceptible effect upon the design conclusions. This trend is demonstrated by the application of these methods to the X-ray (2che[11]) and NMR (1cey #5[15]) structures of CheY which are readily available from the PDB. These structures differ between each other ($rmsd_{BB}$ = 1.9 Å) on the same order as found between the Y13S Spo0F mutant and the refined comparative model of Spo0F. Figure 2 shows the GRID contour maps of the CheY crystal structure and NMR structure resulting from an $N3^+$ probe atom in the active site region. Small differences exist, but both contour maps suggest a common direction for drug discovery. Similarly, the two CheY structures were utilized as target proteins for the DOCK program and evaluated against a portion of our proprietary database. A list of the 300 most favorable ligands from the respective DOCK calculations for the two Spo0F structures (the comparative molecular model vs X-ray structure) and for the two CheY structures (the NMR vs X-ray structures) were generated based upon the potential energy scoring implemented in DOCK and compared. We have found that ~50% of the DOCK hits were found in all lists. If substructure similarity is considered, the percentage of hits common to both structures exceeds 70%, and further, if the dynamic flexibility inherent to the ligands and active site is considered, it is probable that the lists will become nearly identical.

Conclusion

The process of combining comparative modeling techniques with NMR spectroscopy can quickly put three-dimensional structures into the drug discovery effort. Additional structural refinement by extensive experimentation would tend to better define the core region of the protein only since this region is shielded from the ambiguities caused by conformational averaging, and in the case for X-ray structures, crystal packing effects and insufficient side-chain electron density. It is prob-

able that such refinements and adjustments in the core region of the protein would have little effect upon the course of the drug design efforts directed toward the surface, which is properly addressed as dynamic. We believe that the $(\beta\alpha)_5$ fold present in this class of proteins will allow the extension of this method to many of the response regulators, but it should also be considered for other protein classes. Structures produced in this manner require a minimum of experimental data and can allow the drug design process to begin well in advance of the determination of these structures by more exacting experimental means.

Methods

Comparative modeling and protein analysis were performed using MODELER and PROTEIN HEALTH, respectively;[19,21] default values were utilized unless explicitly stated otherwise. The NMR and X-ray structures of CheY were imported from the Brookhaven Protein Data Bank[25] (accession numbers 1CEY and 2CHE). The alignment was conducted manually. Graphical visualization and molecular mechanics minimization were conducted using INSIGHT/DISCOVER.[22] Minimization was conducted using a loop employing 1500 steps of steepest descent minimization followed by 1500 steps of conjugate gradient minimization. Total charge = 0.0, distance dependent dielectric constant = 3.5*r*, cutoff = 21 Å, cutdis = 20 Å, swtdis = 0 Å). The secondary structural restraints were taken directly from the report of Feher et al.[7] In our interpretation, the strong NOEs were given an upper limit of 2.7 Å. Feher et al. did not distinguish between medium and weak NOEs in their study; NOEs reported to be "weak to medium" were given an upper limit of 4.2 Å. The lower limit for all NOE restraints was 1.8 Å; pseudoatom corrections were used for all β-protons, and the glycine α-protons; hydrogen bonds were included only for structured regions of the protein for amide protons in slow exchange ($<10^{-3}$ s^{-1}); for $J_{NH-\alpha H} > 8$ Hz, ϕ was set to $-120 \pm 40°$, and when $J_{NH-\alpha H} < 6$ Hz, ϕ was set to $-60 \pm 30°$, but only for helical regions. In addition to these restraints, based on the crystal structure of CheY, the K104-P105 amide linkage was set to a cisoid configuration ($\omega = 0°$).

The DOCK suite of programs[24] was applied using SYBYL[26] as the graphical interface. For SPHGEN, the maximum and minimum sphere radii are 4.0 and 1.3 Å, respectively, and for CHEMGRID, a 0.4 Å grid size, and a 4*r* distance dependent dielectric were used with a cutoff of 10 Å. Close contact limits were defined as being between 2.3 and 2.8 Å. DOCK was run in the "search" mode with a bin size of 0.5 Å. All other parameters are self-explanatory and are varied according to the number of hits desired and the size of the database to be searched.

Surface properties were determined using the GRID method.[23] All parameters were set to the default values; the grid size was held to 1.0 Å. The GRID data file was converted to a format readable by INSIGHT using the file conversion protocols supplied with the BIOSYM package.

Acknowledgment. We thank Dr. K. I. Varughese for the coordinates of the Y13S Spo0F crystal structure prior to publication. We also acknowledge Drs. Victoria Feher and John Cavanagh for Spo0F NMR data and a preprint of their manuscript. Finally, we thank Gordon J. Madise for his computer systems expertise and support during this project.

References

(1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley and Sons, Inc.: New York, (1986).
(2) The subscript BB refers to the main-chain heavy atoms. All rmsd comparisons were made for residues 3−120.
(3) Clore, G. M.; Gronenborn, A. M. Determination of Structures of Larger Proteins in Solution by Three- and Four-dimensional Heteronuclear Magnetic Resonance Spectroscopy. In *NMR of Proteins*, Clore, G. M.; Gronenborn, A. M., Eds.; The Macmillan Press, Ltd.: London 1993; pp 1−32.
(4) Glusker, J. P.; Trueblood, K. N. *Crystal Structure Analysis: A Primer*; Oxford University Press: New York, 1985.
(5) Olmedo, G.; Gottlin Ninfa, E.; Stock, J.; Youngman, P. Novel Mutations that Alter the Regulation of Sporulation in *Bacillus subtilis*. Evidence that Phosphorylation of Regulatory Protein Spo0A Controls the Initiation of Sporulation. *J. Mol. Biol.* **1990**, *215*, 359−372.
(6) Bourett, R. B.; Borkovich, K. A.; Simon, M. I. Signal Transduction Pathways Involving Protein Phosphorylation in Prokaryotes. *Annu. Rev. Biochem.* **1991,** *60*, 401−441.
(7) Stock, J. B.; Stock, A. M.; Motonen, J. M. Signal Transduction in Bacteria. *Nature* **1990**, *344*, 395−400.
(8) (a) Greer, J. Comparative Model-Building of the Mammalian Serine Proteases. *J. Mol. Biol.* **1981**, *153*, 1027−1042 (b) Greer, J. Model for Haptoglobin Heavy Chain Based Upon Structural Homology. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77 (6)*, 3393−3397.
(9) Volz, K.; Structural Conservation in the CheY Superfamily. *Biochemistry* **1993**, *32*, 11741−53.
(10) Feher, V. A.; Zapf, J. W.; Hoch, J. A.; Dahlquist, F. W.; Whiteley, J. M.; Cavanagh, J. ^{1}H, ^{15}N and ^{13}C Backbone Chemical Shift Assignments, Secondary Structure, and Magnesium Binding Characteristics of the Bacillus subtilis Response Regulator, Spo0F, Determined by Heteronuclear High Resolution NMR. *Protein Sci.* **1995**, *4*, 1801−1814.
(11) Stock, A. M.; Martinez-Hackert, E.; Rasmussen, B. F.; West, A. H.; Stock, J. B.; Ringe, D.; Petsko, G. A. Structure of the Mg^{2+} Bound Form of CheY and Mechanism of Phosphoryl Transfer in Bacterial Chemotaxis. *Biochemistry* **1993**, *32*, 13375−13380.
(12) Volz, K.; Matsumura, P. Crystal Structure of Escherichia Coli CheY Refined at 1.7 Å Resolution, *J. Biol. Chem.* **1991**, 15511−15519.
(13) Stock, A. M.; Motonen, J. M.; Stock, J. B.; Schutt, C. E. Three-dimensional Structure of CheY, the Response Regulator of Bacterial Chemotaxis. *Nature* **1989**, *337*, 745−749.
(14) Santoro, J.; Bruix, M.; Pascual, J.; Lopez, E.; Serrano, L.; Rico, M. Three-dimensional Structure of Chemotactic CheY Protein in Aqueous Solution by Nuclear Magnetic Resonance Methods. *J. Mol. Biol.* **1995**, *247*, 717−725.
(15) Moy, F. J.; Lowry, D. F.; Matsumura, P.; Dahlquist, F. W.; Krywko, J. E.; Domaille, P. J. Assignments, Secondary Structure, Global Fold, and Dynamics of Chemotaxis Y Protein Using Three- and Four-Dimensional Heteronuclear (^{13}C,^{15}N) NMR Spectroscopy. *Biochemistry* **1994,** *33*, 10731−10742.
(16) Bellsolell, L.; Prieto, J.; Serrano, L.; Coll, M. Magnesium Binding to the Bacterial Chemotaxis Protein CheY Results in Large Conformational Changes Involving its Functional Surface. *J. Mol. Biol.* **1994**, *238*, 489−495.
(17) Volkman, B. F.; Nohaile, M. J.; Amy, N. K.; Kustu, S.; Wemmer, D. E. Three-dimensional Solution Structure of the N-Terminal Receiver Domain of NTRC. *Biochemistry* **1995**, *34*, 1413−1424.
(18) The coordinates of Spo0F Y13S mutant were kindly provided to the authors before they were deposited in the Brookhaven Protein Database. See Madhusudan; Zapf, J.; Whiteley, J. M.; Hoch, J. A.; Xuong, N. H.; Varughese, K. I. Crystallization and preliminary x-ray analysis of a Y13S mutant of Spo0F from Bacillus subtilis. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1996**, *D52* (3), 589−590. PDB accession number: 1srr.
(19) MODELER: Molecular Simulations, Inc., 16 New England Executive Park, Burlington, MA 01803−5297.
(20) Sali, A.; Blundell, T. L. Comparative Protein Modeling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.
(21) PROTEIN HEALTH: Molecular Simulations, Inc., 16 New England Executive Park, Burlington, MA 01803−5297.
(22) INSIGHT/DISCOVER: Biosym Technologies, 9685 Scranton Road, San Diego, CA 92121−2777. Discover 2.95 utilizes the CVFF forcefield: Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. Structure and Energetics of Ligand Binding to Proteins: *E coli* dihydrofolate reductase-trimethoprim, a Drug-Receptor System. *Proteins: Struct., Funct. Genet.* **1988**, *4*, 31−47, and references therein.
(23) GRID: Molecular Discovery Ltd., West Way House, Elms Parade, Oxford, OX2 9LL, England, Edition 11. Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Molecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(24) DOCK 3.0 Suite of Programs: Regents of the University of California, 1992. DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D. Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure. *J. Med. Chem.* **1988**, *31*, 722−729.
(25) The Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973.
(26) SYBYL 6.2: Tripos, Inc. 1699 Hanley Rd., St. Louis, MO 63144.

JM970156I