Hypothesis

A model for Fis N-terminus and Fis-invertase recognition

Wen-Shyong Tzou[a], Ming-Jing Hwang[b,*]

[a]*Institute of Life Sciences, National Defense Medical Center, Taipei, Taiwan*
[b]*Institute of Biomedical Sciences, Academia Sinica, 128 Yen-chiou Yuan Rd., Sec. 2, Taipei 11529, Taiwan*

Received 30 September 1996; revised version received 6 November 1996

Abstract In this modeling, we show that the elusive Fis N-terminus involved in invertase-mediated site-specific DNA inversion is, by all indications, a hinged flapping loop. The prediction is based on a combined sequence and secondary structure alignment against known structures of protein segments, as well as a tetrapeptide fragment observed crystallographically. Its validity is strongly supported by the ability to interpret consistently the available mutagenesis data pertaining to this region including, especially, a series of deletion mutants which until this work had been a puzzle in the search for structural explanations. A model for Fis-invertase recognition is also proposed.

Key words: Fis; Invertase; Structure prediction; Homology modeling

1. Introduction

The ability to construct a structural model for invertasome [1] in order to understand the molecular details of its assembly and function has been largely impeded by incomplete knowledge of the structure of a small (~20 amino acids) but functionally important segment in the N-terminal domain of Fis (*F*actor for *i*nversion *s*timulation [2,3]). The current model, supported by mutagenesis data [4,5] and electron microscopic observation of the synaptic formation of invertasome [1], suggests that residues of this small region of Fis interact with the invertase protein, thereby stimulating the activity of DNA inversion [6,7]. Unfortunately, this region is disordered in crystals [8–11] and its structure thus unknown, because its conformation is presumably quite flexible. While attempts have been made to crystallize Fis mutants in the hope that mutation could somehow stabilize the elusive N-terminus and thus allow the observation of its structure, such endeavors have thus far yielded little success [9,10]. Using a computer modeling approach, it is shown here that a flap-hinge configuration emerges as the most probable conformation for the initial 26 amino acid residues of Fis. The resulting model of the Fis N-terminus then allows us to speculate further on a possible Fis-invertase recognition mode. An abstract of this work has appeared elsewhere [12].

2. Materials and methods

The crystal structure of the Fis dimer (PDB [13] code 1fia) is the basis for the present modeling. Amino acid (aa) residues with resolved coordinates are, besides a tetrapeptide of discontinuous electron densities, monomer A: aa 24–98 and monomer B: aa 25–98. All the calculations and molecular structure manipulations were performed with the Discover/InsightII molecular simulation and modeling program and its CFF93 force field (from Molecular Simulation Inc., San Diego, CA; 950 release).

Two methods were applied together to search proteins of known structure in PDB [13] whose segment(s) can serve as a viable structural template for the first 26 amino acid residues of Fis. One employs the BLAST algorithm [14] to search for protein segments in PDB whose sequence is similar to that of the Fis N-terminus (aa 1–26). The other uses the TOPITS algorithm [15] to search for those whose observed secondary structure and solvent accessibility are similar to the predicted values of the Fis N-terminus (also aa 1–26). For the latter, a neural network prediction method (PHD) [16,17] implemented in the TOPITS program was used and its prediction is as follows.

```
.........1.........2......
MPEQRVNSDVLTVSTVNSQDQVTQKP
LEEEEELLLLEEEEEELLLLEEEELL     E: extended     L: loop
obooobobobbbbbbbbobooobooo     b: buried       o: exposed
```

No extensive homologies could be found from either search routine, with p values of all hits from the BLAST search being greater than 0.67, and z scores from the TOPITS search less than 1.61 (a perfect score for p, a probability measure, is zero and $z \geq 3$ is considered to be a good score). To compensate for the low scores, the top 500 hits from both searches were selected and then compared with each other, which yielded 8 common sequences.

The structures of the eight sequences were then superimposed onto the crystal coordinates of Cα atoms for aa 24–26 and a tetrapeptide according to the assignments of Yuan et al. (aa Asp^{20}-Gln^{21}-Val^{22}-Thr^{23}) [10] or Kostrewa et al. (aa Val^{10}-Leu^{11}-Thr^{12}-Val^{13}) [9]. Those with root-mean-square deviation (rmsd) values smaller than 3.0 Å (and where the structure does not bump into the main body of Fis) were selected as a probable structural template for the Fis N-terminus. Each of the final qualified sequences produced by this procedure (there are three) was subsequently mutated into the Fis sequence using a rotamer search algorithm [18] to derive its side chain orientation and, with the superimposed segment, the whole Fis dimer structure was energy-minimized by CFF93, a a second generation (class II) molecular mechanics force field suitable for simulation of organic and biological compounds (for recent references, see [19,20]).

3. A flap-hinge model for Fis N-terminus

As described in Section 2, a search against PDB yielded 8 sequences among the top 500 most homologous to Fis N-terminus (first 26 residues) in both sequence and secondary structure/solvent accessibility, as predicted by the programs BLAST [14] and TOPITS [15], respectively. As might be expected from a previous finding that no significant Fis homologies on both the DNA and amino acid level existed [21], the resulting scores of both searches are low (see Section 2). Nevertheless, these eight peptide segments, whose Cα traces are shown in Fig. 1, provide us with probable conformations to consider for the elusive structure of Fis N-terminus. Furthermore, they can be subjected to additional discrimination based on their superimposed fit on residues 24–26 and a

*Corresponding author. Fax: (886) (2) 785-3569.
E-mail: mjhwang@ibms.sinica.edu.tw

PII S0014-5793(96)01348-8

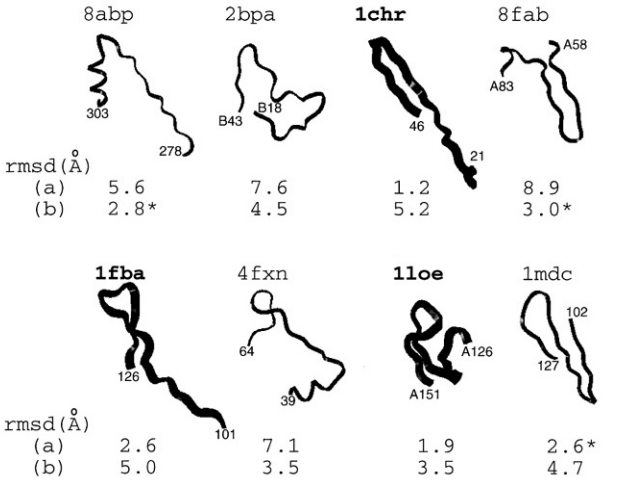

Fig. 1. The eight peptides, identifiable by the labeled sequence numbers and PDB codes of their parent proteins, found to be homologous to the first 26 amino acids of Fis. Their root-mean-square deviations (rmsd) in Cα positions from the X-ray coordinates of residues 24–26 and a tetrapeptide according to the assignment of (a) Kostrewa et al. [9] and (b) Yuan et al. [10] are given. The peptides from 1chr, 1fba, and 1loe (in bold-face) are those having an rmsd under 3 Å with the exclusion of three (*) that have part of the peptide bump the C-terminal domain of Fis (see Fig. 4). The amino acid sequences of the three peptides selected from 1chr, 1fba, and 1loe are given below, along with the corresponding Fis sequence.

Fis	1	MFEQRVNSDVLTVSTVNSQDQVTQKP
1chr	21	MSITTVHQQSYVIVRVYSEGLVGVGE
1fba	101	GIILGIKVDKGVVPLFGSEDEVTTQG
1loe	126	TAWDPSNGDRHIGIDVNSIKSINTKS

tetrapeptide segment, the only residues in this region whose electron densities are clear enough for their coordinates to be resolved [9,10]. The tetrapeptide is peculiar in that it has been assigned to different sequences by the two groups who independently solved the Fis structure [8,10]. Yuan et al. [10] assumed it to be a continuing extension from residue 24, while Kostrewa et al. [9] assigned it to Val^{10}-Leu^{11}-Thr^{12}-Val^{13} based on the argument that the disrupted electron density map matches the characteristic 'Y' shape of leucine side chain and that Leu^{11} is the only leucine residue in this region.

From a structural viewpoint the continuous designation is unlikely to be correct in that: (1) it forces Val^{22}-Thr^{23}-Gln^{24} to adopt a rare inverse γ turn with an unusually short Cα distance between Gln^{21} and Lys^{25} (less than 5 Å which is only half of the normal distance found in typical inverse γ turns [22]); (2) it projects Asp^{20} instead of Leu^{11} into a hydrophobic milieu at a four-way interface involving, in addition to Leu^{11}, the beginning of helix A (Leu^{27}) of the same monomer, the middle portion of helix A of the opposing monomer (Ala^{34}, Leu^{35} and Tyr^{38}), and an early part of helix B (Leu^{53}, Val^{54}, as well as the aliphatic groups of the Glu^{57} side chain), also of the opposing monomer (see Fig. 2); and (3) it would place most of the N-terminal residues away from the large void created in the crystal packing thought to accommodate the disordered structure [9,10].

Nevertheless both assignments were used for the superposition. As seen from Fig. 1, out of the eight peptides, four satisfy a threshold of 3 Å Cα rmsd according to the assignment of Kostrewa et al. [9], of which one possesses several residues running into the main body of Fis. For the continuous designation of Yuan et al. [10], two meet the 3 Å rmsd threshold but both severely bump the DNA-binding domain. This leaves only three segments in the entire PDB [13] whose structure can be considered as a good template for visualizing the disordered Fis N-terminus according to our selection criteria. The consensus conformation resulting from the three segments (bold-face in Fig. 1) as a consequence of the superposition is clearly a flap-hinge type, with each loop segment somehow representing a distinct flapping angle (1chr: small, 1fba: medium, and 1loe: large), as can be seen from Fig. 3. Thus, according to the model depicted in Fig. 4, a short and twisted antiparallel β-sheet formed by residues 11–13 and 24–26 extrudes from the top of the DNA-binding domain to anchor a flapping loop comprised of residues 14–23, while the first 10 amino acids dangle aside.

For the sake of completeness, we continued to examine all other possibilities of assigning the tetrapeptide to a consecutive sequence, but the same three peptide segments remained the only probable structural templates (out of the selected 8) for the Fis N-terminus based on the same criteria of 3 Å cutoff in the superimposed fit and no structural conflict. Within the three peptide segments, several assignments other than Val^{10}-Leu^{11}-Thr^{12}-Val^{13} were also satisfactory. Kostrewa et al. [9] noted that in addition to Leu^{11} they could not rule out the possibility of designating the 'Y' shape electron density to one of the four Asx residues in this region (Asn^7, Asp^9, Asn^{17}, and Asp^{20}). Of the four alternative Y-shape designations, that of Asn^7 and Asp^9 (i.e. assigning the tetrapeptide to aa 6–9 and 8–11, respectively) was able to produce a better or equally good fit than the designation of Leu^{11} in our structural superposition calculations. However, from the discussion given below, the two alternative assignments are clearly not as attractive since neither can result in a structural model that will be reconcilable with the deletion data in this region. Additionally, replacing Leu^{11} by the hydrophilic side chain of Asn^7 or Asp^9 will not be comfortably accommodated in the hydrophobic pocket as noted above for Asp^{20} of the continuous assignment.

The flap-hinge model of Fis N-terminus is capable of being conformationally flexible because the hinges are separated and project upward such that only the loop top is likely to exert a stabilizing interaction with its identical twin of the opposing subunit or with helix A if the loop everts drastically. This is different from most surface loops where they can be stabilized by interacting with surrounding residues. Moreover, the loop

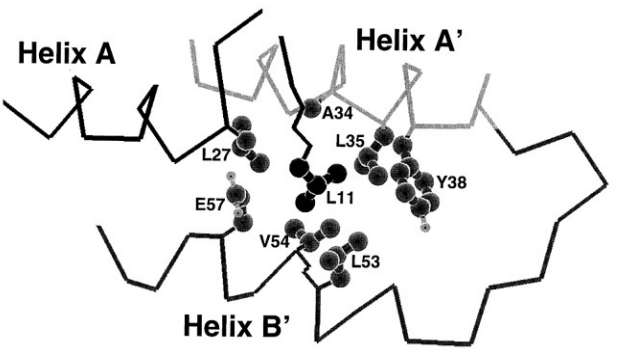

Fig. 2. The hydrophobic milieu at a four-way interface where Leu^{11} is in a pocket surrounded by Leu^{27} of helix A of the same monomer (denoted as helix A), Ala^{34}, Leu^{35}, Tyr^{38} of helix A′, and Leu^{53}, Val^{54}, and Cβ, Cγ, Cδ of Glu^{57} of Helix B′. The pocket is about 10 Å in diameter. Hydrogen atoms have been omitted for the purpose of clarity.

Table 1
Summary of Fis N-terminal mutations and interpretations[a]

Fis mutant	Inversion		Interpretation
	In vitro	In vivo	
(1) Gin-mediated reaction[b]			
Wild type	+	+	
Δ1–5, Δ1–10	+	+	flanking end (1–10) dispensable
Δ1–12	–	–	hinge altered
Δ1–15	–	–	hinge deleted
Δ1–26	–	–	flap and hinge deleted
V16G	–	–	structural integrity of flap affected
V16L/A77V	–	–	structural integrity of flap affected and/or DNA binding ability reduced [5]
K25E	+/–	+/–	a favorable electrostatic interaction with D29 disrupted
K32E	+/–	+/–	change of basic group in A helices slightly affects Fis function [9]
A34V	–	–	bulkier side chain (valine) perturbs the hydrophobic pocket (see text and Fig. 2)
A34V/T23A	–	–	as A34V
(2) Hin-mediated reaction[c]			
Wild type	+	+	
Δ17–21	–	–	flap deleted
Δ18–29	–	–	flap and hinge deleted
Δ24–29	–	–	hinge deleted
F2C	+/–	(–)	1–10 is not required, but point mutation at this region may interfere with the
Q4C	+	+/–	interaction between Fis and invertase
V6I	+	+/–	
V13I	+	+	hinge (β-sheet backbone) not affected
T15I	+	+/–	similar to V16G, but less effect
D20N	(–)	–	direct interaction with invertase involved (see text)
Y29-insert	(–)	(–)	hinge altered (indirectly, see text)
A34P	–	–	break in A helix [9] and/or perturbation of the hydrophobic pocket (see text)
A34T	+/–	+/–	as A34V, but less effect
A34T/G44D	–	(–)	A34T with additional effect from G44D

[a]This table is an updated version of the interpretations presented by Kostrewa et al. [9].
[b]Data from Koch et al. [4]. +, full activity; +/–, weak activity; –, no activity.
[c]Data from Osuna et al. [5]. +, full or strong activity; +/–, weak activity; (–), very weak activity; –, no activity.

so structured is mainly composed of hydrophilic residues including a stretch of five at the loop top (turn), ^{17}Asn-Ser-Gln-Asp-Gln^{21}, consistent with the statistics that loop residues tend to be hydrophilic [23,24] and as such it can be easily solvated. Thus, the three loop structures selected from PDB (Fig. 3) may in fact reflect the conformational dynamics of Fis N-terminus in solution.

4. Interpretation of mutagenesis data

Above all, perhaps the most fingerprinting support of this model is that it effectively satisfies the observation made in a series of deletion experiments where it was shown that while deletion up to the first 10 residues (Δ1–5, Δ1–10) [4] does not affect the stimulating ability of Fis, deletion of further residues (Δ1–12, Δ1–15, Δ1–26 [4] and Δ17–21, Δ18–29, Δ24–29 [5]) abolishes it. Therefore, it would seem that the deleterious effect arises from severe alteration of the anchoring hinge or the flap (e.g. Δ17–21), or both.

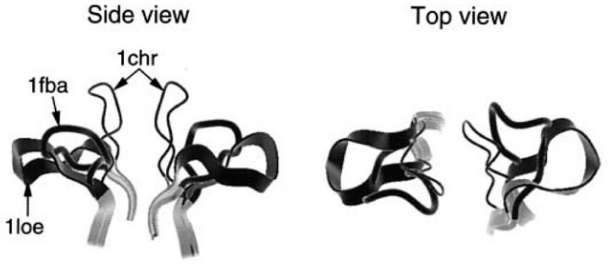

Fig. 3. Two views of the three selected loop structures for Fis N-terminus (aa 1–10 is not displayed for clarity); thin line, 1chr; oval ribbon, 1fba; flat ribbon, 1loe. They are superimposed on two Fis segments (aa 11–13 and 24–26) whose location is indicated by lighter gray.

Besides the deletion data, the model also explains a number of Fis mutants whose enhancer binding is intact but inversion stimulating activity affected [4,5]. Previously, Ala^{34} → Pro was thought to elicit a break in helix A [9], hence a perturbed, thereby functionally defective Fis N-terminus. Our model provides an alternative, which could be more than just additive, source for the N-terminus perturbation. It is the previously unnoted hydrophobic milieu (Fig. 2) which appears to be structurally important as it joins together four parts of the Fis N-terminal domain, including the hinge base. With the present model, the other two mutants at this position, Ala^{34} → Val and Ala^{34} → Thr, can be better interpreted. Valine is much larger than alanine and therefore could significantly perturb the hydrophobic milieu, whereas threonine is intermediate in size and its functional effect can be expected to be, and indeed is, also intermediate. The peculiar Tyr^{29}-insert mutant can be reconciled on the same grounds as well, since even though it may not disrupt helix A, it would place Arg^{28} at the position of Leu^{27} to destroy the hydrophobic joint, and consequently the integrity of the hinge structure. In this model Leu^{11} would be critical because not only is it a hinge residue, but also its side chain projects into the central region of the hydrophobic pocket, apparently serving as a bridge for the other three constituents of the pocket (see Fig. 2). While the effect of the Leu^{11} mutation has not yet been determined experimentally, it is interesting to note that a vicinal mutant,

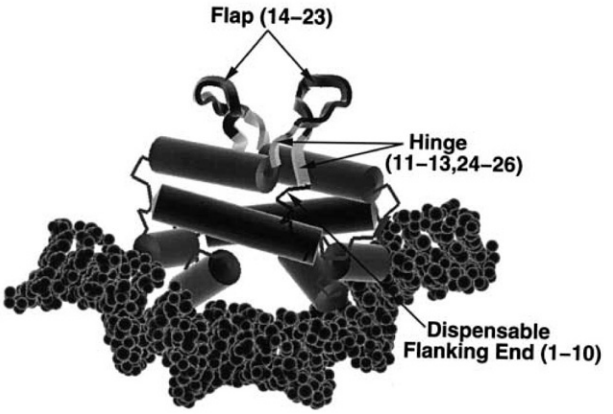

Fig. 4. The emerged flap-hinge model for the crystallographically disordered N-terminal domain of Fis as represented by the peptide segment of 1fba (see Fig. 1) and attached onto the X-ray structure of the helical DNA-binding domain [9,10] with the bound enhancer (a model) [10] shown. Preservation of the Fis dimer dyad symmetry was assumed.

$Val^{13} \rightarrow Ile$, does not affect Hin inversion activity [5]. This ineffective mutation of Val^{13} is yet another piece of data consistent with the present model in that this residue, sitting on top of the hinge, has already moved away from the hydrophobic milieu in space, and such a mutation is unlikely to alter the backbone hydrogen bond of the β-sheet. β-sheet hydrogen bonding is probably also preserved in the mutation of another hinge residue, $Lys^{25} \rightarrow Glu$ [4] of the opposite strand which is located away from the hydrophobic pocket. However, the geometric arrangement of the hinge is such that Lys^{25} can interact with Asp^{29} of helix A to add stability to the hinge structure, therefore the reduced activity observed in $Lys^{25} \rightarrow Glu$ may have been a result of disruption in a favorable electrostatic interaction. As to $Val^{16} \rightarrow Gly$ [4] the interpretation must remain tentative at present; the more flexible and smaller glycine residue may alter the loop structure and/or a critical contact with the invertase. Nevertheless, examination of the predicted loop structures of Fis leads us to speculate that Val^{16} and Val^{22}, the two valines which delimit the five hydrophilic residues presumed to form the loop turn, may interact with each other for the overall integrity of the loop structure. This may help explain, at least partially, the severely impaired activity of the $Val^{16} \rightarrow Leu/Ala^{77} \rightarrow Leu$ double mutant [4]. Based on the flap-hinge model, an updated set of structural interpretations for all the reported Fis N-terminal mutations are summarized in Table 1.

5. A hypothesis for Fis-invertase recognition

Seeing the Fis N-terminus in this appealing model, one is compelled to ask how might it interact with invertase and what are the participating amino acids of both proteins? To address these questions one also needs to know the structure of the invertase catalytic domain which is not yet available. However, a recent experiment [25] strongly suggests Hin forms a dimer resembling that of γδ resolvase [26,27] with which Hin and other members of the invertase family share high (~40%) sequence identity.

Taking the structure of γδ resolvase for invertase, and the predicted flap-hinge model for the Fis N-terminus, we hypothesize that Fis-invertase recognition is facilitated by a dyad axis-to-axis, dimer-to-dimer binding, with residues at the top of the Fis loop interacting with those at the bottom (amino end) of the invertase dimer interface (which is formed between two extensible helices denoted as αE and αE′ according to the nomenclature used for γδ resolvase [26]). The basis for this hypothesis is the following: (1) It was shown that the proposed dimer interface of invertase is critical to the interplay between Fis and Hin; detergents which weaken this interface increase the DNA cleavage rate of Hin [25]. This is consistent with the observation that all the reported Fis-independent invertase mutations [25,28–31] are at or very near this interface. (2) The bottom of this invertase dimer interface is not only symmetry-unique for a 1:1 Fis dimer/invertase dimer recognition, but may also serve as a strategic area of contact for Fis, since perturbation there can be amplified to result in structural alteration at both the dimer interface and the DNA cleaving sites; the latter may be transmitted through a 'scissors-like' motion involving the central β-sheet which had been shown to possess unusual flexibilities [32]. (3) A mapping of invertase sequences to the structure of γδ resolvase shows that located at the proposed Fis-recognition area are three conserved charged residues (Arg^{88}, Asp^{95} and Arg^{102}, Gin sequences; in Hin Arg^{88} is replaced by histidine) whose electrostatic surface may complement that of the hydrophilic residues, ^{17}Asn-Ser-Gln-Asp-Gln^{21}, at the top of the predicted Fis loop. In this context it is interesting to note that mutation $Asp^{20} \rightarrow Asn$ of Fis strongly affects stimulation of Hin-mediated DNA inversion [5].

In summary, in this work we predict the elusive Fis N-terminus to adopt a flap-hinge conformation, and on this basis hypothesize a model for Fis-invertase binding. Whereas the flap-hinge model has been shown to interpret mutagenesis data well, the significance of the proposed Fis-interacting invertase residues requires, and calls for, experimental investigations.

Acknowledgements: This work was supported in part by a grant from National Science Council of Taiwan (NSC85-2311-B-001-030-B16). We wish to acknowledge Drs. H.S. Yuan and R.C. Johnson for stimulating discussions.

References

[1] Heichman, K.A. and Johnson, R.C. (1990) Science 249, 511–517.
[2] Johnson, R.C. and Simon, M.I. (1985) Cell 41, 781–791.
[3] Kahmann, R., Rudt, F., Koch, C. and Mertens, G (1985) Cell 41, 771–780.
[4] Koch, C., Ninnemann, O., Fuss, H. and Kahmann, R. (1991) Nucleic Acids Res. 19, 5915–5922.
[5] Osuna, R., Finkel, S.E. and Johnson, R.C. (1991) EMBO J. 10, 1593–1603.
[6] Finkel, S.E. and Johnson, R.C. (1992) Mol. Microbiol. 6, 3257–3265.
[7] Van de Putte, P. and Goosen, N. (1992) Trends Genet. 8, 457–462.
[8] Kostrewa, D., Granzin, J., Koch, C., Choe, H.-W., Raghunathan, S., Wolf, W., Labahn, J., Kahmann, R. and Saenger, W. (1991) Nature 349, 178–181.
[9] Kostrewa, D., Granzin, J., Stock, D., Choe, H.-W., Labahn, J. and Saenger, W. (1992) J. Mol. Biol. 226, 209–226.
[10] Yuan, S.H., Finkel, S.E., Feng, J.-A., Kaczor-Grzeskowiak, M., Johnson, R.C. and Dickerson, R.E. (1991) Proc. Natl. Acad. Sci. USA 88, 9558–9562.
[11] Yuan, S.H., Wang, S.S., Yang, W.-Z., Finkel, S.E. and Johnson, R.C. (1994) J. Biol. Chem. 269, 28947–28954.
[12] Hwang, M.J. and Tzou, W.S. (1996) Prog. Biophys. Mol. Biol. 65 (Suppl. 1), 139 (Abstract no. P-D1-04, XIIth International Bio-

physics Congress, 11–16 August 1996, Amsterdam, The Netherlands).
[13] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542 (April 1994 Release).
[14] Brenner, S.E. (1995) Trends Genet. 11, 330–331.
[15] Rost, B. (1995) in: The 3rd International Conference on Intelligent Systems for Molecular Biology (ISMB) (Rawlings, C., Clark, D., Altman, R., Hunter, L., Lengauer, T. and Wodak, S. eds.) pp. 314–321, AAAI, Menlo Park, CA.
[16] Rost, B. and Sander, C. (1994) Proteins 19, 55–77.
[17] Rost, B. and Sander, C. (1994) Proteins 20, 216–226.
[18] Mas, M.T., Smith, D.C., Yarmush, D.L., Aisaka, K. and Fine, R.M. (1992) Proteins 14, 483–498.
[19] Maple, J.R., Hwang, M.-J., Stockfisch, T.P., Dinur, U., Waldman, M., Ewig, C.S. and Hagler, A.T. (1994) J. Comput. Chem. 15, 162–182.
[20] Hwang, M.-J., Stockfisch, T.P. and Hagler, A.T. (1994) J. Am. Chem. Soc. 116, 2515–2525.
[21] Koch, C., Vandekerckhove, J. and Kahmann, R. (1988) Proc. Natl. Acad. Sci. USA 85, 4237–4241.
[22] Milner-White, E.J., Ross, B.M., Ismail, R., Belhadj-Mostefa, K. and Poet, R. (1988) J. Mol. Biol. 204, 777–782.
[23] Leszczynski, J.F. and Rose, G.D. (1986) Science 234, 849–855.
[24] Martin, A.C.R., Toda, K., Stirk, H.J. and Thornton, J.M. (1995) Protein Eng. 8, 1093–1101.
[25] Haykinson, M.J., Johnson, L.M., Soong, J. and Johnson, R.C. (1996) Curr. Biol. 6, 163–177.
[26] Sanderson, M.R., Freemont, P.S., Rice, P.A., Goldman, A., Hatfull, G.F., Grindley, N.D.F. and Steitz, T.A. (1990) Cell 63, 1323–1329.
[27] Rice, P.A. and Steitz, T.A. (1994) Structure 2, 371–384.
[28] Haffter, P. and Bickle, T.A. (1988) EMBO J. 7, 3991–3996.
[29] Klippel, A., Cloppenborg, K. and Kahmann, R. (1988) EMBO J. 7, 3983–3989.
[30] Klippel, A., Kanaar, R., Kahmann, R. and Cozzarelli, N.R. (1988) EMBO J. 12, 1047–1057.
[31] Spaeny-Dekking, L., Schlicher, E., Franken, K., Van de Putte, P. and Goosen, N. (1995) J. Bacteriol. 177, 222–228.
[32] Yang, W. and Steitz, T.A. (1995) Cell 82, 193–207.