FEBS Letters 527 (2002) 166–170 FEBS 26447

Desolvation shell of hydrogen bonds in folded proteins, protein complexes and folding pathways

Ariel Fernández[a,b,1,*]

[a] *Institute for Biophysical Dynamics, Cummings Life Science Center #439D, The University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA*
[b] *Institute for Protein Research, Osaka University, Yamadaoka 3-2, Suita, Osaka 565-0871, Japan*

Received 22 May 2002; revised 10 July 2002; accepted 30 July 2002

First published online 15 August 2002

Edited by Gianni Cesareni

Abstract A few backbone hydrogen bonds (HBs) in native protein folds are poorly protected from water attack: their desolvation shell contains an inordinately low number of hydrophobic residues. Thus, an approach by solvent-structuring moieties of a binding partner should contribute significantly to enhance their stability. This effect represents an important factor in the site specificity inherent to protein binding, as inferred from a strong correlation between poorly desolvated HBs and binding sites. The desolvation shells were also examined in a dynamic context: except for a few singular under-protected bonds, the size of desolvation shells is preserved along the folding trajectory.

Key words: Hydrogen bond; Protein binding; Binding interface; Desolvation shell; Local dielectric; All-atom molecular dynamics

1. Introduction

This work deals with a basic question arising as we study protein–ligand binding and protein–protein association [1–4]: is there a particular structural deficiency or anomaly in the native structure of a protein that becomes compensated or gets removed upon binding? In this letter we identify defects in hydrogen-bond packing and show how and why such defects often become determinants of binding sites. In this regard, the basic question that needs to be answered is: what is the influence that the endogenous electrostatic field of a protein molecule exerts on the solvent-structuring moieties – especially overexposed hydrophobes – of its binding partner? These interactions are important in so far as long-range electrostatics are often mediated by solvent and the stability of hydrogen bonds (HBs) – or salt bridges – depends on the polarizability of the surrounding solvent or local dielectric coefficient [5–7].

Let us specialize the study to backbone HBs, the primary determinants of structure. The stabilization of such bonds as they are approached by hydrophobic residues results from the increase in the free energy of the unbound reference state, that is, of the solvent-exposed polar groups involved (amides and carbonyls) [5–7]. The latter increase their solvation free energy if they are exposed to an environment deprived from water or with immobilized or structured water. In this way, there arises a third-body effect, involving the desolvator and the two residues paired by the HB. This effect represents a net force which operates as if the preformed HB were by itself a hydrophobic entity, in agreement with [7].

Thus, a guiding factor in binding results from the enhancement of the endogenous electrostatic field which takes place as poorly desolvated HBs of a single molecule are approached by solvent-structuring moieties of the binding partner, so as to 'complete' their desolvation shells. As shown in this work, this effect becomes an important component in new docking algorithms (cf. [8–12], for instance). At a more fundamental level, it is essentially responsible for the site specificity of binding in those cases where anomalies in hydrogen-bond packing are removed upon association.

2. Materials and methods

In our computations, an amide-carbonyl HB is determined by an N–O distance within the range 2.6–3.5 Å (lower and upper bounds of typical bond lengths) and 0–45° range in the N–H–O angle. The precise location of the amide hydrogen atom cannot be inferred from the crystal structure. Nevertheless, this information is not necessary to obtain the angle between the amide NH and carbonyl CO unit vectors. The identification of hydrophobic residues (L, I, V, F, W, M and A) follows the standard classification based on partition experiments in binary solvents [13].

To introduce a meaningful definition of 'poorly wrapped HB' (PWHB), we first assess the native local solvent environments by defining desolvation spheres of 7.2 Å radius centered at the α-carbons of the residues paired by the HB (our conclusions are fine-tuned and robust only within the range 7.2 ± 0.2 Å). Then, we compute the number of neighboring hydrophobic residues, that is, those whose β-carbon is within the desolvation spheres of the HB. The counting includes any of the residues paired by the HB if they happen to be hydrophobic. Each protective third-body defines a three-body correlation, i.e. a correlation involving the residues paired by the HB and the residue that acts as desolvator of this bond. The total number of such three-body correlations divided by the total number of backbone HBs within the native fold gives a measure, ρ, of the average extent of hydrogen-bond desolvation. This quantity is to be complemented by σ, the dispersion in the extent of hydrogen-bond desolvation within a single molecule.

An inspection of a large PDB sample of native folds reveals that 96% of the 2092 autonomously folded proteins (requiring no chaperone or in vivo environment for folding) of different sizes examined ($33 < N < 400$) have ρ in the range 5.00 ± 0.23. The dispersion σ in the extent of HB wrapping is invariably lower than 19.00%. These basic statistics are illustrated by the examples with structural resolution 2.5 Å or better given in Table 1. Structural redundancies were avoided by intersecting our database with one containing only representative

*Fax: (1)-773-702 0439.
E-mail address: ariel@uchicago.edu (A. Fernández).

[1] On leave from: Instituto de Matemática, UNS-CONICET, Bahía Blanca 8000, Argentina.

0014-5793/02/$22.00
PII: S0014-5793(02)03204-0

proteins and used for protein structure alignment by incremental combinatorial extension of the optimal path [14]. The relatively narrow range of admissible desolvation-sphere radii needed to obtain regular statistics with comparatively low fluctuations implies that the geometric details of native hydrogen-bond packing are fine-tuned to the standard measure adopted to examine them.

The nearly constant ρ value reflects the generic chemical composition of natural chains, while the dispersion σ is essentially due to the wide range of side-chain sizes, which implies that proper desolvation may be achieved by clustering three large residues (like W or F) just as well as clustering six or seven small ones (A).

In this light, an operational definition of PWHB arises focusing on the tails of the distribution of extents of desolvation: a PWHB is one surrounded by at most three hydrophobic residues. Significantly, the N–O distance for all PWHBs and only for those bonds was found to be relatively large, within the range 3.1–3.5 Å.

In the present calculations, the criterion for counting a backbone HB as such requires that its associated interaction energy is less than $-kT$ [5,6]. This definition results in the number of HBs being generally less by about 20% (19% on average) than the number obtained solely using the definition based on a geometric criterion, allowing for a range of 45° in the angle between the amide and carbonyl vectors, and nitrogen–oxygen distance less than 4 Å. The use of the geometric definition without further restrictions has a low quantitative impact on the results shown. For example, the average ratio of the number of three-body correlations versus the number of HBs remained near five (5.31) across our PDB sample. An exception was the dispersion in the average number of three-body correlations per HB, which was about two-fold larger than that reported in Table 1.

A computational tool has been devised to identify the PWHBs for a given PDB structure according to the tenets given above. Thus, sufficiently or poorly desolvated HBs (single or double) of the backbone are represented respectively as thin light grey and thick dark grey segments joining the paired residues at their α-carbons, the backbone is represented as a virtual α-carbon bond black polygonal and hydrophobes are represented as α-carbon spheres, grey if underexposed (>66% buried) and white if overexposed. Fig. 1 displays the backbone HBs for the β-subunit of human hemoglobin: there are only 3 PWHBs: (5,9), strikingly located near Glu6, the residue known to be involved in sickle-cell anemia, and (90,95), (91,95), both at the β-FG corner, a purported binding site in quaternary structure.

Table 1
Number of three-body correlations (C_3), amide-carbonyl HBs (Q), average extent of hydrogen-bond protection ($\rho = C_3/Q$) and dispersion in the extent of protection (σ) for native structures of proteins identified by their PDB accession codes

PDB code	C_3	Q	ρ	σ (%)
1aa2	257	50	5.04	10.18
1lou	242	47	5.15	13.05
1ris	230	45	5.11	12.87
1aue	250	49	5.10	11.80
256b	394	75	5.25	16.05
1abq	80	15	5.33	14.06
1aoj	39	8	4.87	15.47
1ubi	155	31	5.00	10.06
1gb4	80	16	5.00	10.14
1srl	40	8	5.00	12.83
2ptl	74	16	4.62	16.33
1crc	136	28	4.85	9.60
1cw6	32	6	5.33	14.02
1vii	30	6	5.00	12.55
1hhh	446	86	5.18	12.68
1mim	318	64	4.96	17.62
1ifb	215	43	5.00	8.83
1hhg	468	95	4.92	11.09
1e4j	225	45	5.00	12.11
1e4k	233	46	5.07	11.15
1gff-1	612	124	4.93	11.58
1csk-A	111	22	5.04	12.01
1c3t	105	21	5.00	10.78
1a6v	172	33	5.21	17.91

The radius of the desolvation spheres for a HB is fixed at 7.2 Å. Some round-off error in hydrogen-bond counting arises for HBs which are marginally stable and their departure from NH–OC colinearity approaches the critical value of 45°.

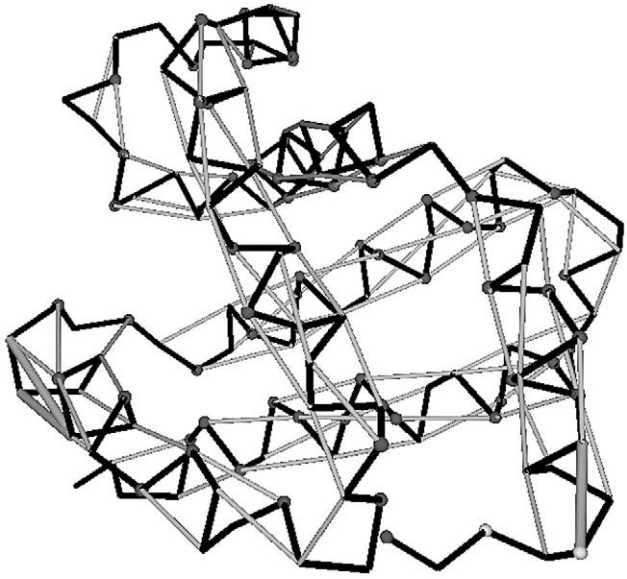

Fig. 1. Pattern of backbone HBs for a β-subunit of human hemoglobin (pdb.1bzo, chain B). The PWHBs are displayed in thick grey segments.

3. Results

3.1. The desolvation shell in protein–protein association

We now address the question: what is the relationship between PWHBs and binding hot spots? Table 2 contains the relevant information extracted from an assortment of high-resolution PDB complexes. The parameters calculated for each complex are: Y_{int}, number of intramolecular PWHBs at binding interface which become sufficiently desolvated ($\rho \leq 3 \rightarrow \rho > 3$) upon binding; Y, total number of PWHBs in both separate partners; $C_{3,\mathrm{int}}$, number of intermolecular three-body correlations (involving an intramolecular (preformed) HB and a hydrophobe from the binding partner); δ, over-all density of PWHBs in both isolated partners; and δ_{int}, PWHB density at binding interface. The densities were calculated as number of PWHBs per 1000 Å^2, and the solvent-exposed surface areas of complexes and separate units were computed using the program GetArea 1.1 [15].

We invariably found that 80–100% of PWHBs at the interface became sufficiently wrapped upon binding ($\rho > 3$) and that the density of PWHBs is significantly larger – in some cases as much as seven times larger – at the interface when compared with the over-all average density of the separated binding partners. This fact implies that the three-body correlations represent an important factor to guide the binding process and that binding specificity is to a considerable extent the result of the thermodynamic benefit entailed by the desolvation of PWHBs: the PWHBs are primary determinants of binding hot spots.

Notwithstanding the strength of this signal, not every PWHB in a given protein can be rationalized in terms of a binding site.

As an illustration, the structural information on the dimeric HIV-1 protease pdb.1a30 complex [16] is displayed in Fig.

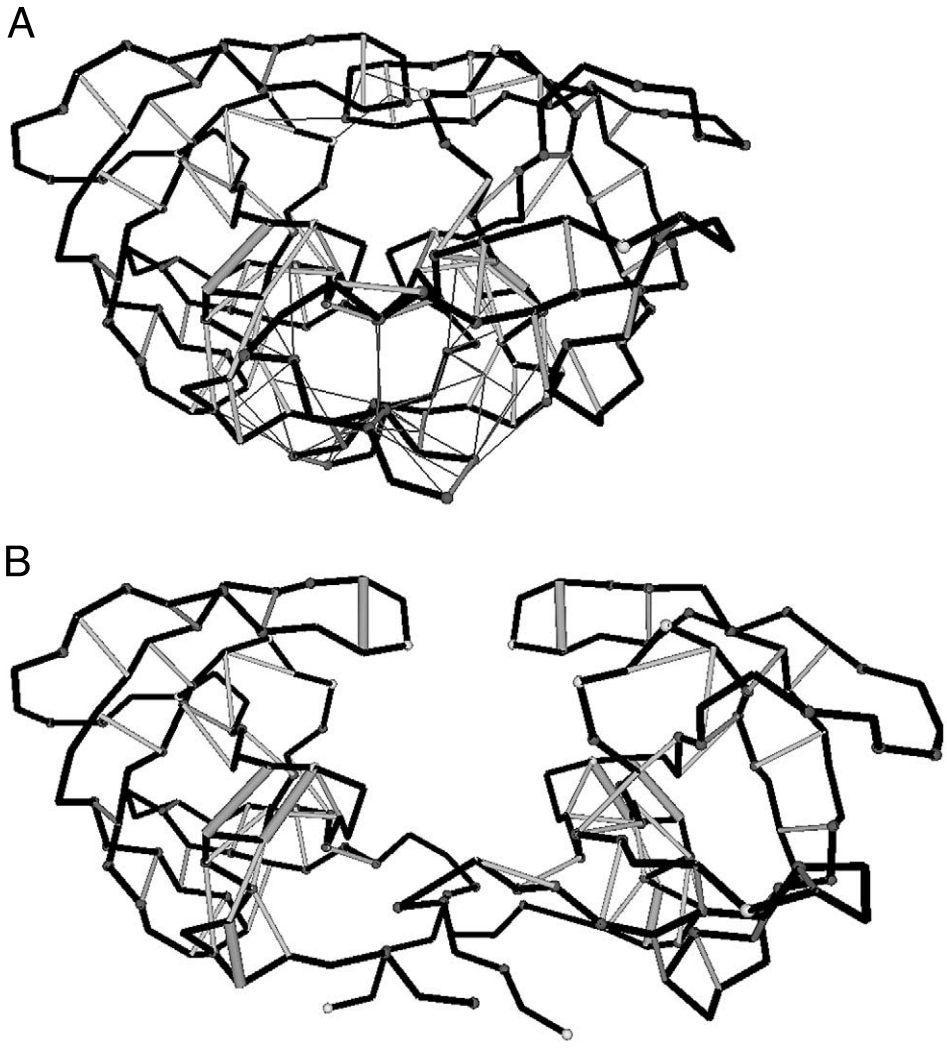

Fig. 2. A: Dimeric HIV-1 protease complex (pdb.1a30, chains A, B) displaying the intermolecular three-body correlations. B: The separated monomers, revealing the intramolecular PWHBs whose desolvation shell is completed upon association as a result of the intermolecular three-body correlations.

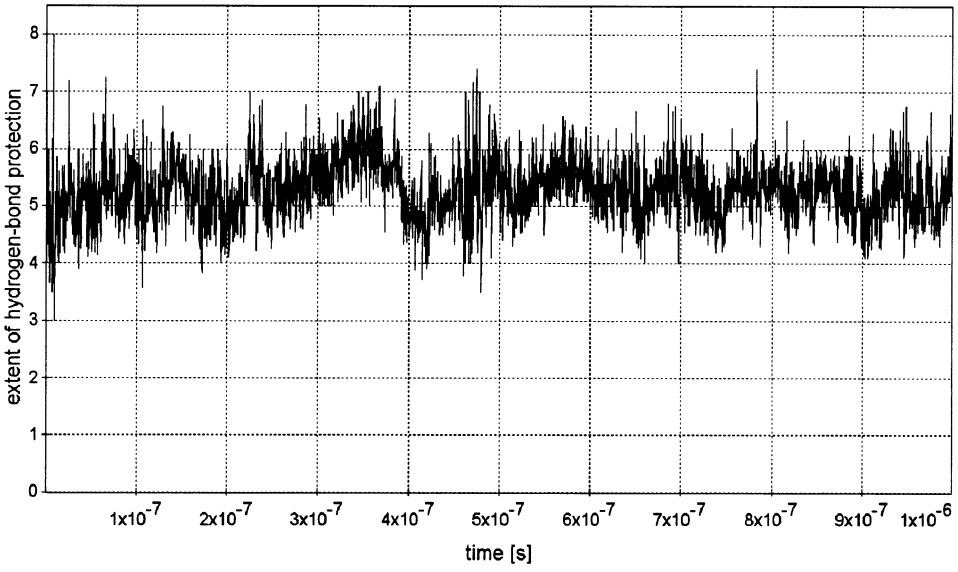

Fig. 3. Time-dependent average extent of protection of backbone HBs, $\rho = \rho(t)$, for the longest available all-atom folding trajectory [17].

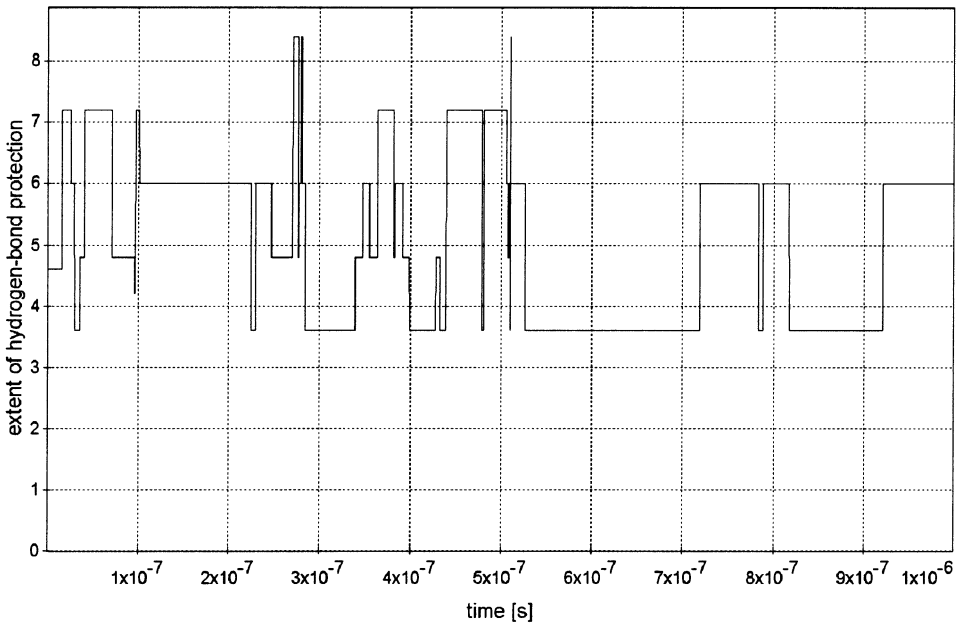

Fig. 4. Time-dependent average extent of protection of backbone HBs, $\rho = \rho(t)$, obtained independently from a coarse computation of villin headpiece folding with implicit solvent (cf. [5,6]).

2A,B. The complex with its intermolecular three-body correlations (thin lines from the protecting or desolvating hydrophobes to the center of the HB) is displayed in Fig. 2A. By contrasting it with the isolated units (Fig. 2B), we can identify the intramolecular single or double PWHBs which are stabilized upon complex formation, i.e. those whose desolvation shells are completed upon association.

3.2. Are desolvation shells preserved along folding pathways?

We are now in a position to systematically examine HB desolvation shells along folding pathways and address the question: what is the dynamic relevance of the HB desolvation shells? To investigate this question, we re-examined the longest all-atom trajectory available, the Duan–Kollman MD trajectory with explicit solvent [17], which simulates 1 μs of the folding time (~5 μs) of the villin headpiece ($n = 35$). The average extent of desolvation is now time-dependent, $\rho = \rho(t)$, and displayed in Fig. 3. Its Gaussian dispersion, $\sigma = \sigma(t)$ over all HBs occurring at each given time, is invariably under 18.82%. Fig. 3 and the upper bound on σ reveal that the average extent of hydrogen-bond desolvation $\rho = 5$ is very nearly a constant of motion for the folding trajectory. This

Table 2
Structural information extracted from an assortment of high-resolution PDB complexes

Complex name – PDB Code	Y_{int}	Y	$C_{3,int}$	δ (10^{-3} Å^{-2})	δ_{int} (10^{-3} Å^{-2})
Insulin – 6ins	6	7	29	0.80	4.51
HIV-1 protease+inhibitor – 1a30	21	26	107	1.87	4.71
SIV protease – 1siv	9	14	60	1.06	2.65
Defensin – 1dfn	9	14	21	2.72	10.01
Barstar+mutant – 1a19	4	18	28	1.80	3.91
Subtilisin+eglin-C – 1cse	15	21	54	1.51	9.70
Antitrypsin polyms. – 1d5s	14	22	176	1.01	2.76
Bombyxin – 1bon	4	5	18	0.60	3.02
FcγRIII – 1e4k, B–C	7	22	19	0.97	7.08
SH3+ligand – 1prl	1	3	3	0.55	1.45
Colicin+ligand – 1bxi	6	12	19	0.92	3.97
Colicin+ligand – 1emv	5	11	20	0.86	3.20
SH3 novel dimer – 1aoj	18	20	143	1.72	4.46
Serpin+ligand – 1as4	14	31	169	1.40	2.02
Anti-oncogene – 1a1u	5	13	51	1.84	2.45
Troponin heterodimer – 1pon	6	10	25	1.34	4.54
MHC, antigen+receptor – 1im9, A–D	3	22	18	0.84	2.22
MHC, antigen+ligand – 1im9, A–C	3	19	20	2.00	6.12
Insulin+ligand – 1cph	8	12	55	2.43	4.80
De-novo protein of α-2D – 1qp6	8	12	56	1.65	3.67
Spectrin – 2spc	25	56	154	2.98	4.86

The parameters calculated for each complex are: Y_{int}, number of PWHBs at binding interface which become sufficiently desolvated ($\rho > 3$) upon binding; Y, total number of PWHBs in both individual partners; $C_{3,int}$, number of intermolecular three-body correlations (involving an intramolecular HB and a hydrophobe from the binding partner); δ, over-all density of PWHBs in both isolated partners; and δ_{int}, PWHB density at binding interface. The densities were calculated as number of PWHBs per 1000 Å^2.

observation suggests that the same HB-building constraints present in native folds govern the entire folding process.

This result was further corroborated by independent ab-initio simulations of the villin headpiece folding based on a judicious coarsening of the backbone torsional dynamics. The algorithm used to generate the reproducible trajectories is described in detail in [5,6] and is based on the following premises: (a) inter-basin hopping within a single Ramachandran map is incommensurably slower than intra-basin exploration and thus, the backbone (Φ, Ψ)-torsional dynamics may be described by a coarse stochastic process defined by the time evolution of Ramachandran basin assignments to the residues; (b) the probability for a given residue to change its Ramachandran basin depends on its extent of structural involvement, i.e. the more engaged the residue is, the less prone to undergo a basin hopping; (c) the pairwise contributions to the intramolecular energy are rescaled at each iteration according to the environment surrounding each pairwise interaction, in turn determined by the three-body correlations.

Such simulations, each consisting of 10^6 iterations, were performed under the same conditions as the all-atom trajectory and were reproducible in 10 out of 14 runs. The reproducible runs invariably reveal an extent of hydrogen-bond protection virtually identical to the all-atom trajectory, as shown in Fig. 4.

The pervasiveness of the $\rho = 5$ desolvation shell observed in both all-atom and coarse computations implies that proteins do not make HBs incrementally: most HBs are either considerably desolvated from their inception or not present at all, and those few who remain under-desolvated play a decisive role in protein–protein association.

4. Discussion

This work emphasized the generic importance of structural defects, specifically poorly wrapped HBs (PWHBs), as factors determining binding sites in protein–protein interactions. Furthermore, we have delineated their singularity in the dynamics of the protein folding process.

The significantly larger density of PWHBs at the protein binding interfaces and the fact that most of such preformed PWHBs become fully desolvated upon association support the view that a considerable thermodynamic benefit is achieved by completing their desolvation shells. Moreover, a complete desolvation shell for a HB appears to be a building constraint that applies along the entire folding trajectory, not simply to native structures.

Acknowledgements: The author thanks Prof. Y. Duan for making the Duan–Kollman trajectory available for analysis, and Profs. Robert Huber, R. Stephen Berry, Stuart Rice, Tobin R. Sosnick, Yuji Goto, Tao Pan, Karl Freed and Philippe Cluzel for enlightening discussions.

References

[1] Ma, B., Shatsky, M., Wolfson, H. and Nussinov, R. (2002) Protein Sci. 11, 184–197.
[2] Clackson, T. and Wells, J.A. (1995) Science 267, 383–386.
[3] Bogan, A.A. and Thorn, K.S. (1998) J. Mol. Biol. 280, 1–9.
[4] Ringe, D. (1995) Curr. Opin. Struct. Biol. 5, 825–829.
[5] Fernández, A. (2002) Proteins Struct. Funct. Genet. 47, 447–457.
[6] Fernández, A. (2001) J. Chem. Phys. 114, 2489–2502.
[7] Roseman, M.A. (1988) J. Mol. Biol. 201, 621–624.
[8] Norel, R., Sheinerman, F., Petrey, D. and Honig, B. (2001) Protein Sci. 10, 2147–2161.
[9] Bliznyuk, A.I. and Gready, J.E. (1999) J. Comput. Chem. 20, 983–992.
[10] Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesman, A., Aflalo, C. and Vakser, I. (1992) Proc. Natl. Acad. Sci. USA 89, 2195–2199.
[11] Connolly, M.L. (1983) Science 221, 709–711.
[12] Hendsch, Z.S. and Tidor, B. (1999) Protein Sci. 8, 1381–1392.
[13] West, M.W., Wang, W., Patterson, J., Mancias, J., Beasley, J.R. and Hecht, M.H. (1999) Proc. Natl. Acad. Sci. USA 96, 11211–11216.
[14] Shindyalov, I.N. and Bourne, P.E. (1998) Protein Eng. 11, 739–747.
[15] Fraczkiewicz, R. and Braun, W. (1998) J. Comp. Chem. 19, 319–333.
[16] Louis, J.M., Dyda, F., Nashed, N.T., Kimmel, A.R. and Davies, D.R. (1998) Biochemistry 37, 2105–2112.
[17] Duan, Y. and Kollman, P.A. (1998) Science 282, 740–744.