



## NMR-restrained docking of a peptidic inhibitor to the N-terminal domain of the phosphoenolpyruvate:sugar phosphotransferase enzyme I

Didier Rognan<sup>a,\*</sup>, Seema Mukhija<sup>b</sup>, Gerd Folkers<sup>a</sup> & Oliver Zerbe<sup>a</sup>

<sup>a</sup>*Department of Applied Biosciences, Swiss Federal Institute of Technology, Zürich, Switzerland;* <sup>b</sup>*ARPIDA AG, Münchenstein, Switzerland*

Received 28 December 1999; Accepted 31 July 2000

**Key words:** docking, Enzyme I, nuclear magnetic resonance, phosphotransferase, peptide

### Summary

Starting from the NMR structure of the binary complex between the N-terminal domain of the unphosphorylated enzyme I (EIN) of the phosphoenolpyruvate:sugar phosphotransferase (PTS) and the histidine-containing phosphocarrier protein (HPr), a molecular model of the phosphorylated transition state of the related complex was established using constrained simulated annealing. The coordinates of the phosphorylated EIN enzyme were then used in a second step for flexible docking of a decapeptide inhibitor of EIN whose enzyme-bound conformation itself was determined by NMR using transferred nuclear Overhauser effects. Two phosphorylation models of the peptide inhibitor were investigated and shown to be both functional. Interestingly, one model is very similar to that of the complex between EIN and its natural substrate HPr. The present study demonstrates that NMR-guided flexible docking constitutes an interesting tool for docking highly flexible peptide ligands and facilitates the upcoming protein-based design of nonpeptide EIN inhibitors for discovering new antibiotics.

### Introduction

The phosphoenolpyruvate-phosphotransferase system (PTS) [1] consists of a cascade of several enzymes responsible for the concomitant transport and phosphorylation of carbohydrates in bacteria [2, 3]. In response to the availability of sugars, it also modules catabolite repression, gene expression and chemotaxis. As the PTS is unique to the prokaryotic kingdom, certain enzymes of the PTS are potential targets for the discovery of new anti-infectives that could overcome the ever increasing resistance of bacterial pathogens to known antibiotics [4].

Enzyme I (EI) is the first enzyme of the PTS cascade which, after autophosphorylation of its His189 by phosphoenolpyruvate (PEP), transfers its phosphoryl group to His-15 of the Histidine phosphocarrier protein (HPr). In the subsequent step HPr phosphorylates a second class of enzymes, collectively named

Enzymes II (EII) which are responsible for sugar phosphorylation and translocation (for a review see [3]). We focussed our attention on the EI-HPr complex for several reasons: (i) each enzyme of the binary complex is structurally well determined both before [5–8] and after phosphorylation [9–12], (ii) the recently solved NMR structure of the N-terminal domain of EI (EIN) in complex with HPr [13] can be used as a starting point for a protein-based inhibitor design program, (iii) short peptides were found which inhibit EIN and which are phosphorylated by EI at a histidine residue found in all inhibitory peptides [14].

Unfortunately, no coordinates of EIN-inhibitor complexes are currently available. In order to gain insight into the possible binding modes of EIN inhibiting compounds, known inhibitors have first to be docked into the active site of EIN. Docking of small molecular-weight ligands to a macromolecule can be done using either matching or docking simulation techniques. Matching [15, 16] and incremental docking [17] procedures are usually fast and allow to dock entire chemical databases however at the price

\*To whom correspondence should be addressed; E-mail: didier@pharma.ethz.ch

of a rather crude estimate of the interaction energy. In addition, possible conformations of the molecules to be docked need to be pre-selected among a reduced list of putative dihedral angles. More detailed data of the energetics during a flexible docking require docking simulations using either Monte Carlo-Simulated Annealing [18, 19], genetic algorithms [20, 21] or evolutionary programs [22]. Although genetic algorithms seem to cope better with ligand flexibility [21], docking of highly flexible ligands (up to 20 rotatable bonds) is still computationally very challenging. To restrain the overall conformational space of the protein-bound ligand, additional experimental constraints are necessary.

An elegant way to solve this problem is to use structural information about the bound ligands. Although coordinates of ligand-protein complexes for tight binding ligands may principally be completely derived from nuclear magnetic resonance (NMR) spectroscopy [23–25], this method is very time-consuming or even impossible for large protein targets. To speed up the structure calculation, isotope-editing techniques have been employed [24]. However, these require either the receptor or the ligand to be isotopically enriched which may be difficult or impossible for the latter if it is non-peptidic or contains non-natural amino acids. The transfer NOEs (tr-NOEs) technique [26, 27] has successfully been used to derive the structure of ligands when bound to their protein targets. In this experiment, a 10–100 fold excess of ligand is mixed with the target protein. NOE build-up during the NOESY mixing time takes place mainly during the time the ligand is bound but is read-out on the free form with its favourable NMR properties (high concentration, sharp lines). This NMR experiment is easily performed on non-labelled protein and ligand but requires that the binding and dissociation range of the enzyme-ligand complex are within the limited range compatible with tr-NOE measurements. Furthermore, the ligand should ideally display no or very weak NOEs in the free form. Provided that these conditions are met (equilibrium dissociation constant  $K_D$  in the 10 mM–10  $\mu$ M range, see [28]) and that the protein is purified in sufficient amounts (to about 50  $\mu$ M in 250–500  $\mu$ l) structural information on the bound conformation of the ligand may be derived rather quickly. It is important to note that even if the molecular weight of the protein-ligand complex is too large to observe protein resonances directly, tr-NOEs can still provide very useful infor-

mation to determine the conformation of the bound ligand (for recent reviews see [29, 30]).

In the present study, we propose a general strategy for docking a highly flexible decapeptide ligand to an enzyme active site. Using structural information derived from NMR, we docked into the active site of EIN a decapeptide which had been shown to inhibit EI and to be phosphorylated by EI at His5 [14]. The obtained three-dimensional (3D) models of the enzyme-inhibitor complex are discussed in the light of known experimental data and constitute a first step in the rational design of selective EIN inhibitors.

## Material and methods

### *Construction of plasmid pMSEINH6*

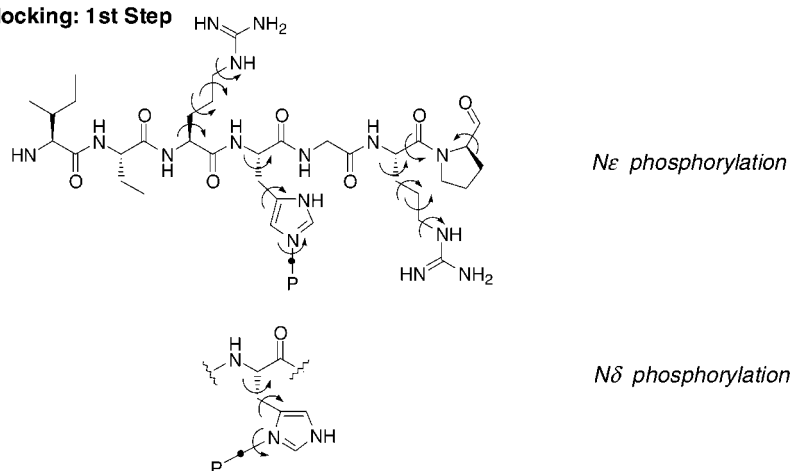
Plasmid pMSEINH6 for the controlled expression and purification by Ni-NTA chromatography of the N-terminal domain of EI was constructed as follows: The coding region was PCR amplified with forward primer 5'-GGTAGGCATATGATTTTCAGG-3', reverse primer 5'-GGGGGAAGCTTTTTAGTGATGGTGATGGTGATGTTTAGCATGCTCTGCTTTTCAGAAGCC-3' and plasmid pTSHIC9 [31] as template. The PCR product was digested with NdeI and HindIII, and the 788 bp fragment was ligated with the vector fragment of pMS470d8 (a derivative of pJF119EH) [32] obtained by digestion with the same restriction enzymes. Standard procedures were used for plasmid purification, restriction analysis, ligation and transformation [33].

### *Expression and Purification of Recombinant EINH6*

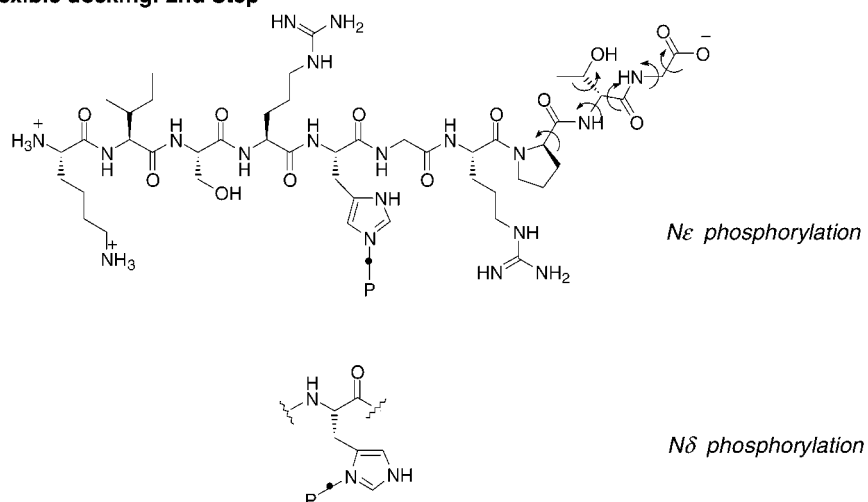
*E. coli* ZSC112L [34] transformed with plasmid pMSEINH6 was grown in LB medium to  $OD_{600} = 0.8$ , induced with 100  $\mu$ M isopropyl-thio- $\beta$ -D galactopyranoside and allowed to grow overnight. Cells from 1 l culture were harvested by centrifugation, resuspended in buffer A (20 mM TrisHCl pH 7.5, 1 mM EDTA, 150 mM NaCl, 0.2 mM  $\beta$ -mercaptoethanol), ruptured in a French pressure cell, and the cytoplasmic fraction was freed of membranes and insoluble material by centrifugation. The supernatant was adsorbed Ni-NTA agarose (20 ml, Qiagen) [35] equilibrated with buffer A, the resin was washed with 200 ml buffer A containing 25 mM imidazole, and EIN was then eluted with buffer A containing 250 mM imidazole. The eluate was concentrated and further purified on Superdex 75 (Pharmacia) equilibrated with buffer A.

Decapeptide 1: **KISRHRG**RPTG

**Flexible docking: 1st Step**



**Flexible docking: 2nd Step**



*Figure 1.* Structure of peptide **1**, and definition of free rotating bonds assigned for the two-steps docking to the phosphorylated form of EIN. The heptapeptide core used for the first docking step is indicated in bold face. A dummy atom (dot) is inserted between His5 Nδ (or Nε) atom and the central phosphorus atom, in order to allow global rotation of the imidazole ring around the P-N axis.

*NMR determination of the conformation of peptide 1 bound to unphosphorylated EIN*

The sample for measurement was prepared by adding purified peptide **1** (Figure 1) (>99% by RP-HPLC) to an aqueous solution (9:1 H<sub>2</sub>O/D<sub>2</sub>O v/v) of EIN dissolved in buffer. The buffer contained 20 mM NaPi, 150 mM NaCl, 0.5 mM DTT, 1 mM EDTA at pH 7.5. The final concentration of the peptide was 2 mM

making up a peptide:protein concentration ratio of 30:1. Preliminary test measurements indicated that the pH was too high to yield reasonably sharp amide resonances and was therefore brought up to pH = 5.7 by adding small aliquots of HCl. The enzymatic activity was only slightly decreased at pH 5.7 as checked in a phosphoenolpyruvate:sugar phosphotransferase assay (data not shown).

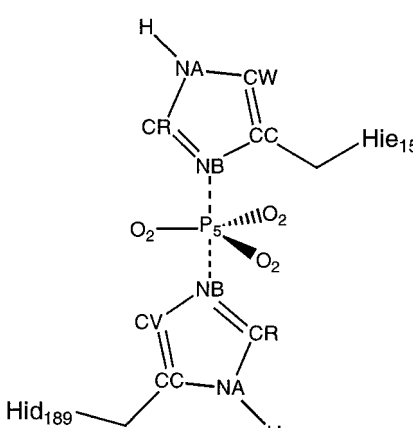
The tr-NOEs data were derived from a 200 ms NOESY experiment recorded at 295 K on a Bruker DRX-600 spectrometer accumulating 64 scans per increment. The final data matrix contained  $2048 \times 512$  complex data points. Water suppression was achieved through a watergate-type element prior to detection in order to minimize saturation effects of amide resonances from solvent suppression. Spin-diffusion effects were largely eliminated by including only those NOEs that were also present in a 50 ms NOESY. The sign of the NOEs derived from the peptide resonances changed from positive in the absence of EIN to negative in the peptide/protein mixture verifying that correct conditions for observation of tr-NOEs had been selected.

The spectra were processed using Bruker XWIN-NMR software and transferred to XEASY [36] which was used for data interpretation and integration of NOEs. Assignments for the peptide resonances were performed on the free peptide, largely following the sequential resonance assignment protocol developed by Wüthrich [37]. Peak positions were only slightly shifted in the peptide:protein mixture. The peak intensities were then translated into distance constraints and structure calculations by restrained molecular dynamics were performed within the DYANA package [38] using the implemented standard simulating annealing protocol in dihedral angle space. No stereospecific assignments could be obtained at the resolution obtained from the data. The DYANA target function that reflects how much the NOEs are satisfied by the resulting structures was reasonably low ( $<1.5$ ) [38] and no consistent violations were observed in the final assignment. However, the calibration had to be loosened slightly in order to account for the increased flexibility inherent to the method. The resulting 20 lowest-energy structures were superimposed for best fit of the backbone atoms in order to identify well-structured regions of the peptide.

#### Parameterisation of the phosphoryl group

The phosphoryl group was first parameterised for the AMBER5.0 programme [39] using the AMBER95 force-field [40] starting from a pentacoordinated phosphorus atom using a trigonal bipyramidal geometry with both histidine nitrogen atoms in apical positions 2.0 Å away from the central phosphorus atom [41] (see Table 1 for force-field parameters). Atomic charges for the phosphoryl group were calculated using the GAUSSIAN94 package [42] and the HF/6-31G\* basis

Table 1. AMBER5.0 parameters for the phosphoryl group.



Atoms types		
Atom <sup>a</sup>	Atom type	Mass
phosphorus	P5	30.97
oxygen	O2	16.00
Bond Parameters		
Bonds	$r_{eq}^b$	$K_r^c$
NB-P5	2.000	525
O2-P5	1.480	525
Angle Parameters		
Angles	$\Theta_{eq}^d$	$K_\Theta^e$
NB-P5-NB	180.0	140
NB-P5-O2	90.0	140
O2-P5-O2	120.0	140
CV-NB-P5	120.0	140
CC-NB-P5	120.0	140
CR-NB-P5	120.0	140
Van der Waals parameters		
Atom type	$R^{*f}$	$\epsilon^g$
P5	2.1000	0.20
O2	1.6612	0.21
Atomic charges		
Atom	Charge <sup>h</sup>	
P5	0.967	
O2	-0.989	

<sup>a</sup>Atom types are assigned as previously described [39] except for the phosphoryl phosphorus atom for which a new atom type (P5) has been created.

<sup>b</sup>Equilibrium bond length, in Å.

<sup>c</sup>Harmonic force constant, in kcal/(mol.Å<sup>2</sup>).

<sup>d</sup>Equilibrium angle bend, in degrees.

<sup>e</sup>Harmonic force constant, in kcal/(mol rad<sup>2</sup>).

<sup>f</sup>Van der Waals radius, in Å.

<sup>g</sup>6-12 potential well depth, in kcal/mol.

<sup>h</sup>Restrained Electrostatic Potential atomic charge.

set. Therein, atom-centred charges were fitted to an *ab initio* electrostatic potential using the RESP method [43] under a previously-described protocol [44].

#### *Building a three-dimensional model for the phosphorylated form of Enzyme I*

The solution structure of the EIN-HPr complex [13] (PDB entry: 3eza) was used as a template for deriving a model for the phosphorylated enzyme. The phosphorus atom of a phosphoryl group was placed at the centre of a dummy bond linking the two complexing histidine nitrogen atoms (N $\epsilon$  of His189<sub>E</sub> for EIN, N $\delta$  of His15<sub>H</sub> for HPr) followed by deletion of the initial dummy bond. Covalent bonds between the central phosphorus atoms and both histidine nitrogen atoms were then defined in the LINK module of the AMBER5.0 package [39]. The phosphoryl transfer EIN-HPr complex was further minimized in two steps using the above-described trigonal-bipyramidal coordination geometry of the phosphorous as a constraint. First, all residues located in a 10 Å-radius sphere centred on the phosphoryl group were submitted to a 10-ps Simulated Annealing (SA) protocol. Starting with velocities randomly assigned to match a Boltzman distribution at a temperature of 1000 K, the complex was first coupled for 5 ps to a heat bath at 1000 K using a temperature coupling constant  $\tau_T$  of 0.2 ps and then linearly cooled down to 50 K for the next 5 ps. During these 10 ps, all atoms located outside the above-defined sphere were kept fixed at their initial NMR coordinates. Since the simulated annealing was performed *in vacuo*, a distance-dependent dielectric function ( $\epsilon = 4r$ ) was used. A twin cut-off (10.0, 15.0 Å) was used to calculate non-bonded electrostatic interactions at every minimization step and to update every non-bonded pair list at every 10 steps, respectively. The last SA conformer was then fully minimized without any restraint by 1000 steps of steepest descent followed by conjugate gradients relaxation until the rms gradient of the potential energy was less than 0.01 kcal/mol Å. Removing the HPr atoms from the minimized structure afforded the coordinates of the phosphorylated EIN enzyme (EINP) used for further docking purpose.

#### *NMR-restrained automated docking of the peptide inhibitor to the phosphorylated N-terminal domain of Enzyme I*

The automated flexible docking of the NMR-determined enzyme-bound conformation of decapep-

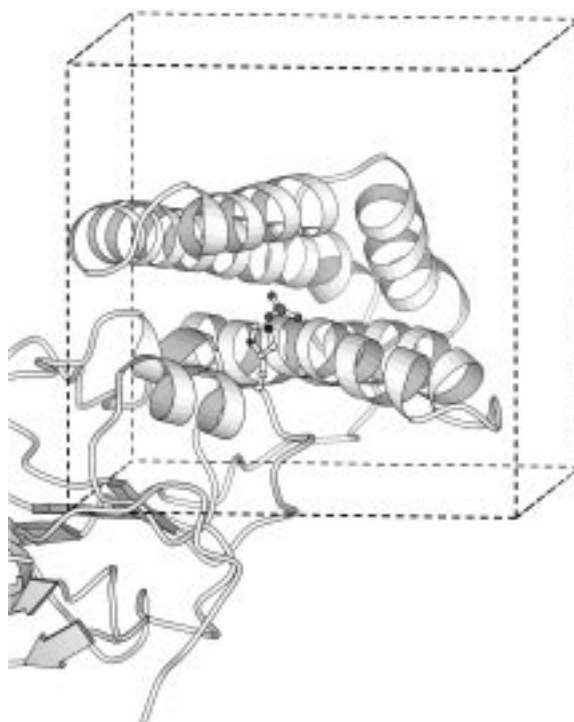
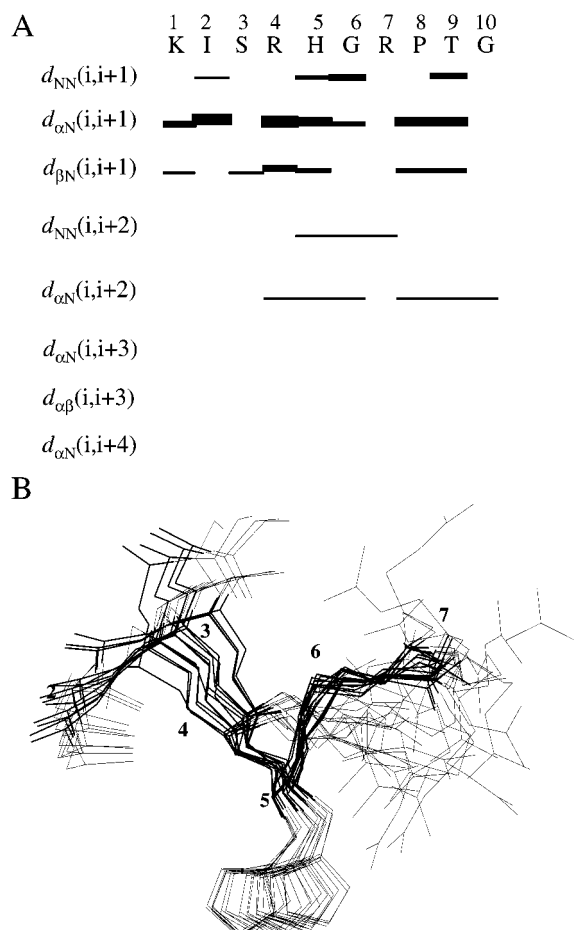


Figure 2. Three-dimensional box encompassing the phosphorylated EIN active site. The orientation of the phosphoryl group with respect to His189<sub>E</sub> is indicated. The figure has been done using Molscript v.2.1 [47].

tide **1** into the binding site of the phosphorylated EIN was performed using the AutoDock 2.4 programme [19] which utilises a combined Monte-Carlo (MC)/Simulated Annealing (SA) docking approach. In order to limit the number of freely rotating bonds, the docking was realized in two steps: (i) the central I<sup>2</sup>SRHGRP<sup>7</sup> heptapeptide segment containing residues that are well-defined by the previously described tr-NOEs was docked first. During the docking, the conformation of both the backbone and of well-defined side chains was kept constant and only bonds of ill-defined side chains were allowed to rotate; (ii) starting from conformation derived in step (i), the heptapeptide was extended to the full-length decapeptide **1**. In the subsequent docking the initial heptapeptide core was frozen and all new single bonds were allowed to rotate. Because peptide **1** is known to be phosphorylated by EI at His5, it was orientated such that the phosphoryl group of EINP was coordinated with either the N $\delta$  or the N $\epsilon$  His5 atom of the peptide, thereby restricting the initial random orientations of the peptide in the three-dimensional lattice during docking.



**Figure 3.** NMR measurement of E1N-bound peptide **1**. (a) Plot displaying the distribution of specific NOE's over the sequence. The thickness of the lines is proportional to the volume of the corresponding peak. The distances are selected to distinguish the secondary structural elements. The number of sequential (medium-range) constraints used in the DYANA calculation is 20(3) for residue 2, 18(1) for residue 3, 10(3) for residue 4, 13(1) for residue 5, 13(3) for residue 6 and 17(1) for residue 7.

(b) 20 nmr structures of E1N-bound peptide, fitted over backbone atoms from residues 2 to 7. Ca positions are labeled from 2 to 7. The figure has been prepared using Molmol [48].

Docking of the Heptapeptide I<sup>2</sup>SRHGRP<sup>7</sup> core: First, a three-dimensional grid (size: 33.8 Å × 33.8 Å × 33.8 Å, resolution: 0.375 Å) was centred on the phosphoryl group of E1NP (Figure 2). Steric and electrostatic interaction energies between the enzyme and several probe atoms (C, N, O, H) representing the different atom types of the peptide ligand were pre-computed before the docking phase for each point of the three-dimensional lattice. AMBER 4.1 non-bonded parameters and atomic charges were used for that purpose. A sigmoidal distance-

dependent dielectric function [45] was used for calculating electrostatic interactions. The two initial orientations of the heptapeptide were defined such that either the Nδ or the Nε His5 atom was placed in an apical position, 2 Å away from the central phosphorus atom of the phosphoryl moiety. In order to keep the likely trigonal bipyramidal coordination of the phosphorus atom, no global rotation or translation was allowed. Rotation about the P-Nδ(ε) coordination bond was possible by defining a dummy atom, which was not taken into account during the force-field calculations (Figure 1). The dummy atom was defined to lie within the coordination sphere of the phosphoryl P atom of E1NP. 50 independent docking runs of 50 cooling cycles each (from 500 to 40 K) were established starting from the orientation of the peptide in the three-dimensional grid as indicated above. Within each SA cycle, freely rotating bonds were varied randomly according to a MC protocol. Once 3000 MC steps were either accepted or rejected according to the classical Metropolis criterion [46], the Boltzmann temperature factor was decreased and another cycle of MC docking performed starting from the most energetically favoured state of the previous cycle. Thereby, at least 50 × 50 × 3000 (7,500,000) bound states had been scanned by the flexible docking procedure. The most favoured bound state for each independent docking run was saved, and the resulting ensemble of structures examined by a cluster analysis based on rms deviations from the starting orientation of the bound ligand. Two different orientations were clustered in the same family whenever rms deviations of all ligand atoms were below 1.0 Å. The energetically most favourable state for each orientation of peptide bound to the phosphoryl group via the Nδ or Nε atom of His5 was selected for the second docking step.

Docking the full decapeptide **1**: Starting from the preferred orientation of the heptapeptide core found in the first docking step described above, the peptide was extended at both termini by adding the N-terminal lysine and the C-terminal Thr-Gly dipeptide in a fully extended conformation. Both terminal residues were charged. The decapeptide **1** was then docked in the binding site of E1NP using the previously defined grid and a similar docking protocol. Freely rotating bonds were limited to the C-terminal adduct (Figure 1) since the N-terminal lysine could by no means contact the phosphorylated enzyme (see justification in the Results section).

Table 2. Automated docking of the ISRHGRP heptapeptide.

Cluster <sup>a</sup>	Nδ phosphorylation model			Nε phosphorylation model		
	E <sub>min</sub> <sup>b</sup>	E <sub>mean</sub> <sup>c</sup>	Orientations <sup>d</sup>	E <sub>min</sub>	E <sub>mean</sub>	Orientations
1	-60.44	-55.16	10	-59.03		1
2	-48.70	-46.92	4	-50.89		1
3	-48.18	-40.41	27	-43.06		1
4	-38.75		1	-39.60	-35.10	2
5	-37.67	-34.26	5	-37.06		1
6	-34.55	-32.76	2	-33.13	-28.87	5
7	-32.90		1	-32.51		1
8				-31.90	-27.24	2
9				-31.56		1
10				-30.09	-27.18	3
11				-29.22		1
12				-27.86		1
13				-27.04		1
14				-26.14	-23.55	6
15				-26.12	-24.69	3
16				-25.47		1
17				-25.33		1
18				-24.80		1
19				-24.10		1
20				-23.84		1
21				-23.65		1
22				-22.73		1
23				-22.37	-21.58	9
24				-22.37		1
25				-21.05		1
26				-20.18		1
27				-18.31		1

<sup>a</sup>Cluster number, classified by increasing enzyme-peptide interaction energy.<sup>b</sup>Lowest interaction energy (kcal/mol) within a cluster. When not specified, E<sub>min</sub> = E<sub>mean</sub>.<sup>c</sup>Mean interaction energy (kcal/mol) for all complexes of the same cluster: Intermolecular interaction energy + peptide internal energy.<sup>d</sup>Number of similar orientations in the cluster.

The two-steps docking procedure required about 7 and 2 hours for the heptamer and decamer, respectively, on a R10K INDIGO SGI workstation.

## Results and discussion

### *NMR determination of the Enzyme I-bound conformation of peptide 1*

We were able to show that data from tr-NOEs measurements (Table 2) provided enough constraints to define the binding part of the decapeptide **1** with sufficient precision to enable modelling of the peptide:protein complex. In a superposition of the 20 lowest-energy structures from the NMR-restrained

MD calculation, residues 3–7 exhibited root mean square deviations (rmsd) values below 1 Å whereas the terminal backbone atoms and the side chains were much less well-defined (Figure 3). It is therefore likely that only the central part of the peptide is involved in binding. Residues 3–7 form a tight (β) turn although the usually observed i,i+3 hydrogen bond cannot be found in the ensemble of NMR structures. This is not necessarily to be expected since the structure is more likely stabilized through interactions with protein residues. Due to the limited number of constraints and the lack of stereospecific assignments of side chain methylene protons, most side chains remain largely disordered. However, a comparison of the rmsd values for side chains 3–8 with the residual ones

(side chains 1–2, 9–10) reveals restricted spatial freedom for the side chains of the turn-forming residues. Considering that side chains are usually less well-defined in NMR structures and that the tr-NOEs data deliver less precise structural data, it seems to be very likely that these side chains are directly contacting the protein. Their limited spatial resolution unfortunately required most of them to remain flexible during the docking calculation (*vide infra*).

### *Three-dimensional model of the phosphorylated form of EINP*

The decapeptide **1** is phosphorylated by EINP [14]. As phosphorylation of EIN locally changes the environment of His189 [12, 13], we felt that the phosphorylated form of EIN (EINP) would be better suited for peptide docking than the non-phosphorylated form. Furthermore, a model for the phosphorylated enzyme, ready to accommodate the peptidic inhibitor, had to be built first since no experimentally determined structure is available for phosphorylated EINP. Starting from the recently published NMR structure of the binary complex between EIN and its natural substrate HPr, a phosphoryl group was placed between both molecules according to the experimental data: (1) the phosphoryl group was located between the two histidine side chains that participate in the phosphoryl transfer, His189<sub>E</sub> (His189 of Enzyme I) and His15<sub>H</sub> (His15 of HPr), (2) a trigonal bipyramidal coordination of the phosphorus atom was chosen with both histidine nitrogen atoms (Nε2 of His189<sub>E</sub>, Nδ1 of His15<sub>H</sub>) located in axial positions and the three phosphoryl oxygen atoms in equatorial positions (see force field parameters of the phosphoryl transition state in Table 1). We are aware of the fact that the phosphoryl group should polarize the two histidines (His189<sub>E</sub>, His15<sub>H</sub>) that coordinate the phosphorus atom. This was not taken into account in the parameterisation (Table 1) of the phosphoryl transition state, because it was topologically restrained (accurate definition of covalent interactions) and because the two coordinating histidines did not interact with other residues in our modeled EINP-HPr complex. Thus, inadequate charges on the two histidines should not alter non-bonded interactions significantly. After refinement of the starting structure by simulated annealing and energy minimisation, a model of the phosphorylated complex could be readily obtained (Figure 4b). The side chain of His189<sub>E</sub>, which is one of the two phosphorus-complexing histidines, switched from a g+ ( $\chi^1 =$

178°,  $\chi^2 = 78^\circ$ ) to a g- ( $\chi^1 = -148^\circ$ ,  $\chi^2 = -118^\circ$ ) rotameric state. It should be noted that the rotameric state of the HPr histidine (His15<sub>H</sub>) was not affected by the refinement. In the proposed model, the phosphoryl oxygen atoms are hydrogen-bonded to Lys69<sub>E</sub> and Thr16<sub>H</sub> side chains. Thereby, the phosphoryl group is in an ideal position to be transferred from EIN to HPr.

### *NMR-restrained docking of peptide 1 to EINP*

The NMR structure of the peptide was docked in two steps : (1) docking of the averaged heptapeptide NMR structure with a rigid backbone and flexible side chains, (2) extending the docked conformation of the heptapeptide to the full-length decapeptide while maintaining the geometry of the whole heptapeptide core and allowing conformational flexibility only to the newly built C-terminal end.

Since the backbone conformation of EIN-bound peptide **1** was rather well conserved from Ile2 to Arg7 within the ensemble of NMR structures (Figure 3), we initially docked the averaged coordinates of the heptapeptide ISHRGRP core, assuming a rigid backbone (excepted for the C-terminal proline) and flexible side chains for Arg4, His5 and Arg7 (Figure 1). This docking strategy is supported by two experimental data: (i) Both arginine side chains were not particularly well resolved in the NMR study (Figure 3) so that no rotameric state could be favoured for these two basic residues, (ii) although the His5 side chain seems to adopt a preferential conformation when bound to the unphosphorylated EIN (Figure 3) recent NMR studies indicate that phosphorylation of EIN alters the rotameric state of the phosphorylated histidine [9, 12, 13]. Therefore, the His5 side chain was allowed to be flexible during the first docking phase. Since no experimental data were available to favour one phosphorylation site over the other, both nitrogen atoms (Nδ1 and Nε2) were considered as potential phosphorylating atoms (Figure 1) leading to two possible sets of docking solutions. To ensure optimal coordination, the histidine nitrogen atoms were first placed 2 Å away from the central phosphorus at the apex of a putative trigonal bipyramidal and the coordinates of the Nδ (or Nε) atoms of His5 were fixed, thereby preventing translational diffusion of the peptide. The heptapeptide was then docked by searching the accessible conformational space. During that process no overall rotational/translational diffusion was permitted in order to fix the coordinates of the His5 Nδ1 (or Nε2) atom. A dummy atom was inserted into to



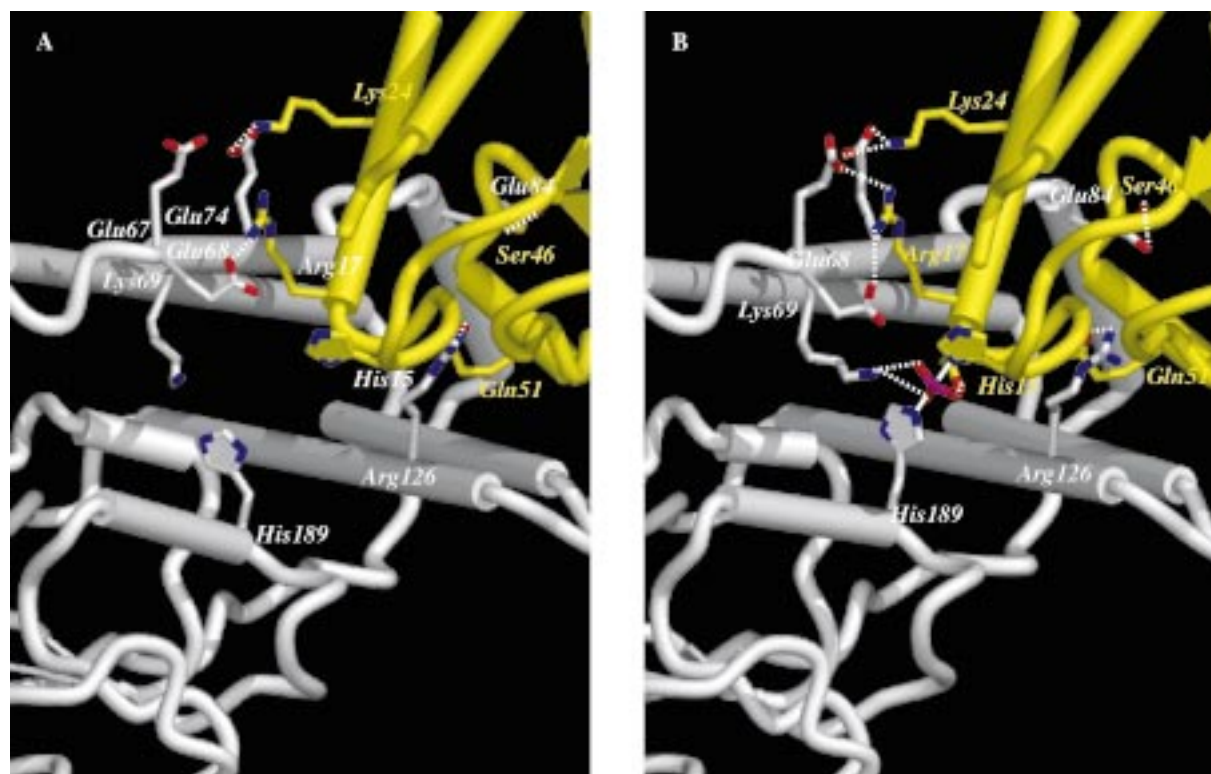


Figure 4. Close-up to the interface between EIN and HPr (a) unphosphorylated NMR structure; (b) phosphorylated three-dimensional model. EIN are HPr are displayed in white and yellow, respectively. H-bonds are indicated by white dashed lines. Coordination of the central P atom by the two histidine nitrogens is displayed by a solid white line.

the phosphorus-N bond to allow for rotation of the imidazole ring around the N-Du bond (Figure 1).

Interestingly, much fewer solutions were found for the N $\delta$  phosphorylation model (total of 7 different clusters) than for the N $\epsilon$  model (27 possible orientations, see Table 3) although the first-ranked solutions were energetically similar ( $-60.44$  kcal/mol for the N $\delta$  model vs.  $-59.03$  kcal/mol for the N $\epsilon$  model). Using a 20 kcal/mol threshold, 3 orientations were finally selected from the first model and 4 from the second one (Figure 5).

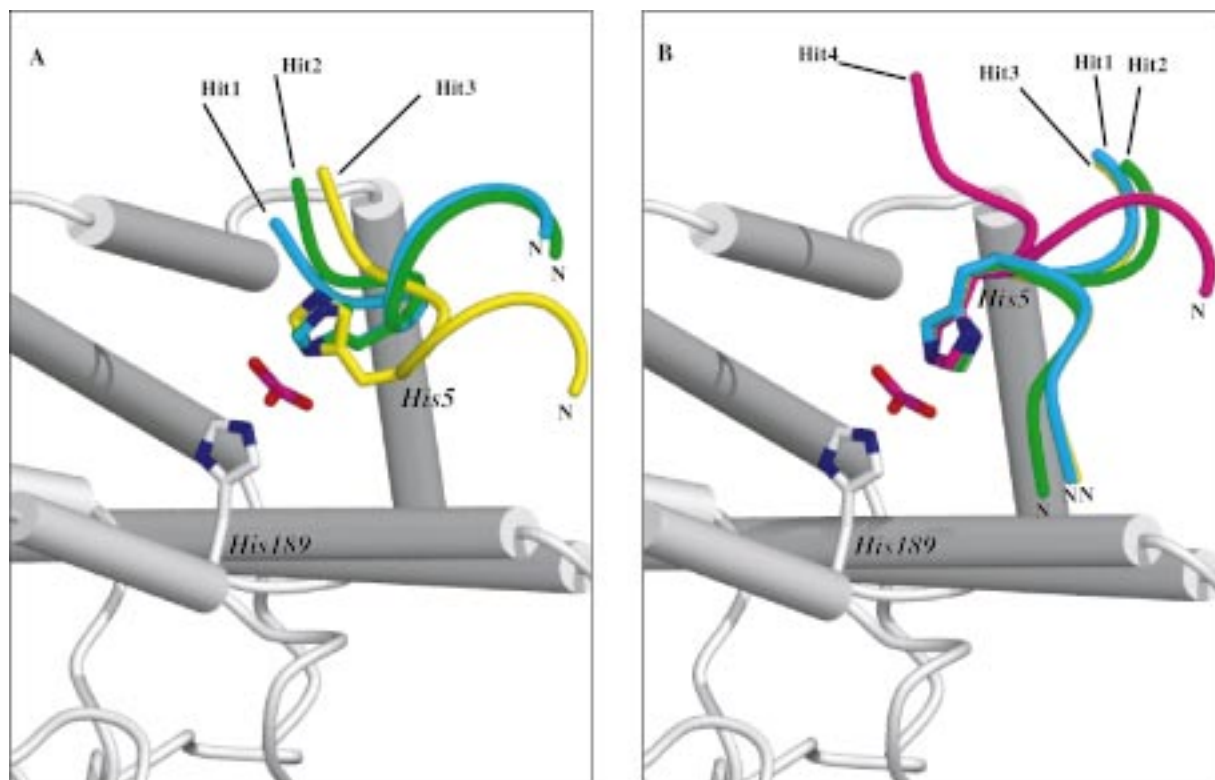
For the N $\delta$  model, the first-ranked orientation is energetically favoured over the two others (Table 3). Moreover, in this orientation the C-terminal residue (Pro8) is optimally positioned for backbone extension (Figure 5a). In the two other solutions, Pro8 comes so close to the enzyme surface that major conformational changes would have been necessary in order to accommodate the two additional amino acids that have to be added to yield the full-size peptide (Thr9 and Gly10). For the N $\epsilon$  model, the first free-docking solutions are very similar with respect to the overall

orientation of the peptide in the enzyme cleft. The fourth docked orientation (Figure 5b) is different but energetically disfavoured (interaction energy difference of 19.43 kcal/mol). Therefore, the top-ranked solution of this model was also chosen for further protein-restricted ligand binding.

Completion of peptide building to the final decapeptide **1** was achieved by adding an N-terminal lysine and a C-terminal Thr-Gly dipeptide. Lys1 was added in a trans orientation to the previously selected enzyme-bound heptapeptide structures for both phosphorylation models. This conformation was chosen arbitrarily since Lys1 cannot contact any enzyme atoms whatever its conformation. The best bound conformation of the C-terminal part was searched by a second flexible docking step during which the orientation of the core heptapeptide was maintained (Figure 1). Two solutions were found for the N $\delta$  model and the most frequently populated orientation was finally selected (Table 4). For the N $\epsilon$  model, all 50 docking runs converged towards a unique solution for protein-based flexible ligand building (Table 4).

Table 3. Automated docking of the Decapeptide **1**.

Rank	N $\delta$ phosphorylation model			N $\epsilon$ phosphorylation model		
	E <sub>min</sub>	E <sub>mean</sub>	Orientations	E <sub>min</sub>	E <sub>mean</sub>	Orientations
1	-73.24	-72.15	39	-69.45	-67.99	50
2	-72.07	-71.57	11			

Figure 5. Preferred orientation of the heptapeptide bound to EIN (a) N $\delta$  phosphorylation model; (b) N $\epsilon$  phosphorylation model.

#### *Hypothetical EINP-bound conformation of peptide **1***

In the N $\delta$  phosphorylation model, 7 H-bonds exist between the peptide and the enzyme in the proposed transition state. In addition to His5, for which the location of the N $\delta$  atom was constrained during the docking (see experimental procedure), Arg7 is the main anchoring residue with its side chain H-bonded to the phosphoryl oxygen atoms and to Asp129 (Figure 6a). Two additional H-bonds are found between peptide backbone atoms (amide oxygen of Arg4 and Gly6) and neighbouring enzyme side chains (Tyr122 and Arg126). Interestingly, the second basic side chain (Arg7) is orientated towards two conserved negatively-charged residues of EIN (Glu67 and Glu68), exactly like the Arg17 side chain of the nat-

ural substrate HPr (compare Figures 4b and 6b). A salt-bridge to the Glu68 side-chain is formed. The N-terminal end of the peptide was pointing out of the binding site in accordance to the experimental tr-NOE data that indicate that these residues are flexible. The C-terminal end folds back towards the enzyme after residue Thr9. Due to the highly polar character of the peptide ligand there are rather few lipophilic interactions. The only noticeable one involves Pro8 whose ring is facing the Ile72 side chain.

In the N $\epsilon$  phosphorylation model, another orientation of the peptide has been selected in which both basic side-chains (Arg4 and Arg7) are only interacting with the phosphoryl group. The N-terminal end (Lys1-Ser3) is completely expelled from the binding

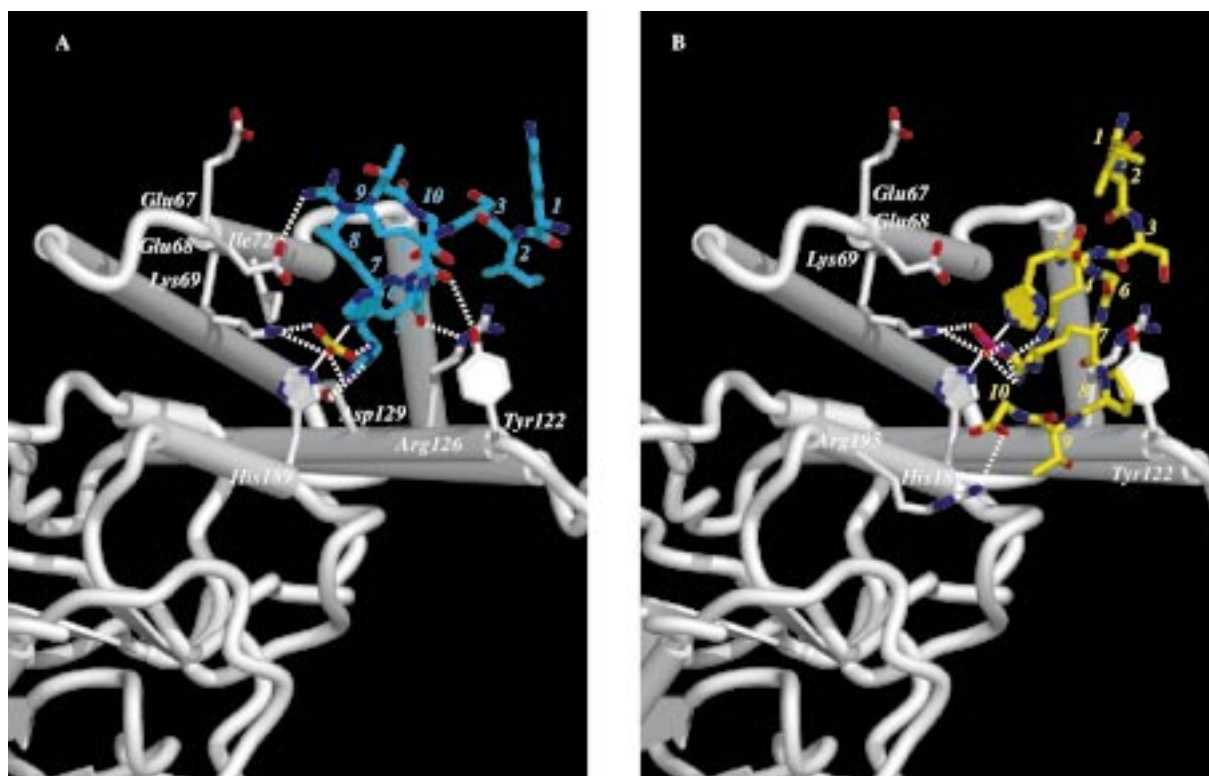


Figure 6. Preferred orientation of the decapeptide **1** bound to EIN. (a) N $\delta$  phosphorylation model; (b) N $\epsilon$  phosphorylation model. H-bonds are indicated by white dashed lines. Coordination of the central P atom by the two histidine nitrogens is displayed by a solid white line. Figures 4–6 have been done using the PREPI software [49] and rendered with PovRay [50].

site whereas the C-terminal carboxylate is interacting with the basic Arg195 side chain. Again Pro8 provides the only lipophilic contacts to the Tyr122 aromatic ring.

From an energetic point of view, it is difficult to favour one of the two proposed models. Intra- and intermolecular interaction energies are rather similar for both orientations (Table 3). However, it must be noted that fewer orientations for the N $\delta$  model were found for the core heptapeptide in the first docking round (Table 2). Furthermore, the most favoured cluster was higher populated and of lower intramolecular energy in that model. If we assume that this heptapeptide core provides the main interactions to the enzyme (a statement that is well-supported by the tr-NOEs NMR data) the latter model should be favoured. In agreement with this hypothesis we also point out that the first interaction model is less dependent on the phosphoryl group than the N $\epsilon$  model (2 intermolecular H-bonds vs. 4). As the peptide **1** also binds to the nonphosphorylated enzyme, the N $\delta$  model should be clearly favoured in that context. Site-directed mutagenesis of

EIN residues predicted to specifically interact with the decapeptide in both models (Glu68, Ile72, Asp120, Arg126 for the N $\delta$  model; Arg195 for the N $\epsilon$  model) should provide an experimental support for one of the two postulated hypotheses.

## Conclusion

The conformation of an oligopeptide bound to the N-terminal domain of EIN has been determined by  $^1\text{H}$ -NMR using the transferred NOEs technique. The proposed conformation was used to provide restraints for a flexible docking of the peptide to the active site of EIN. The NMR-determined backbone conformation of the peptide was kept fixed while anchoring side chains were allowed to rotate freely during a two-step docking procedure: (i) docking of a core heptapeptide for which the backbone conformation was well resolved by NMR, (ii) docking of the full decapeptide, starting from the previously proposed orientation and bound conformation of the core heptapeptide. Interestingly, the NMR-determined backbone conformation

of the peptide fitted nicely into the enzyme-binding site and some modifications of the rotational states of only a few anchoring side-chains were necessary. Two models of peptide phosphorylation were investigated and shown to be energetically equivalent. However, one model for which the N $\delta$  atom of a peptidic histidine residue is phosphorylated upon binding is favoured in the light of experimental binding data. The present model has already been used for protein-based lead identification of  $\mu$ M non-peptidic inhibitors using virtual screening of a chemical database (unpublished data). More generally, it shows that NMR-restrained docking is a valuable tool for docking highly flexible ligands.

### Acknowledgements

The authors wish to thank Dr Gillesen (ARPIDA AG, Munchenstein) for financial support, Professor B. Erni (University of Bern, Switzerland) for critical reading of the manuscript and the Computing Centre of the ETHZ for allocation of computing time on the CRAY-J90 cluster and the Intel Paragon machines.

### References

- Kundig, W., Ghosh, S. and Roseman, S., *Proc. Natl. Acad. Sci. USA*, 52 (1964) 1067.
- Postma, P.W. and Lengeler, J.W., *Microbiol. Rev.*, 49 (1985) 232.
- Herzberg, O. and Klevit, R., *Curr. Opin. Struct. Biol.*, 4 (1994) 814.
- Setti, E.L. and Micevitch, R.G., *Curr. Med. Chem.*, 5 (1988) 101.
- Herzberg, O., Reddy, P., Sutrina, S., Saier, M.H., Jr., Reizer, J. and Kapadia, G., *Proc. Natl. Acad. Sci. USA*, 89 (1992) 2499.
- Jia, Z., Quail, J.W., Waygood, E.B. and Delbaere, L.T.J., *J. Biol. Chem.*, 268 (1993) 22490.
- Liao, D.I., Silverton, E., Seok, Y.J., Lee, B.R., Peterkofsky, A. and Davies, D.R., *Structure*, 4 (1996) 861.
- Garrett, D.S., Seok, Y.J., Peterkofsky, A., Clore, G.M. and Gronenborn, A.M., *Biochemistry*, 36 (1997) 4393.
- Rajagopal, P., Waygood, E.B. and Klevit, R.E., *Biochemistry*, 33 (1994) 15271.
- van Nuland, N.A., Boelens, R., Scheek, R.M. and Robillard, G.T., *J. Mol. Biol.*, 246 (1995) 180.
- Van Nuland, N.A., Wiersma, J.A., Van Der Spoel, D., De Groot, B.L., Scheek, R.M. and Robillard, G.T., *Protein Sci.*, 5 (1996) 442.
- Garrett, D.S., Seok, Y.J., Peterkofsky, A., Clore, G.M. and Gronenborn, A.M., *Protein Sci.*, 7 (1998) 789.
- Garrett, D.S., Seok, Y.J., Peterkofsky, A., Gronenborn, A.M. and Clore, G.M., *Nat. Struct. Biol.*, 6 (1999) 166.
- Mukhija, S. and Erni, B., *Mol. Microbiol.*, 25 (1997) 1159.
- Gwschend, D.A. and Kuntz, I.D., *J. Comput. Aid. Mol. Des.*, 10 (1996) 123.
- Ewing, T.J.A. and Kuntz, I.D., *J. Comput. Chem.*, 18 (1997) 1175.
- Rarey, M., Kramer, B., Lengauer, T., Klebe, G., *J. Mol. Biol.*, 261 (1996) 470.
- Morris, G.M., Goodsell, D.S., Huey, R. and Olson, A., *J. Comput. Aid. Mol. Des.*, 10 (1996) 293.
- Goodsell, D.S. and Olson, A., *J. Proteins*, 8 (1990) 195.
- Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R., *J. Mol. Biol.*, 267 (1997) 727.
- Morris, G.M., Goodsell, D.S., Halliday, R., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J., *J. Comput. Chem.*, 19 (1998) 1639.
- Gehlhaar, D.K., Verkhivker, G.M., Rejto, P.A., Sherman, C.J., Fogel, D.B., Fogel, L.J. and Freer, S.T., *Chemistry & Biology*, 2 (1995) 317.
- Gronenborn, A.M. and Clore, G.M., *Crit. Rev. Biochem. Mol. Biol.*, 30 (1995) 351.
- Petros, A.M. and Fesik, S.W., *Meth. Enzymol.*, 239 (1994) 717.
- Wand, A.J. and Short, J.H., *Meth. Enzymol.*, 239 (1994) 700.
- Clore, G.M. and Gronenborn, A.M., *J. Magn. Reson.*, 53 (1983) 423.
- Anglister, J. and Naider, F., *Methods Enzymol.*, 202 (1991) 228.
- Campbell, A.P. and Sykes, B.D., *J. Magn. Reson.*, 93 (1991) 77.
- Moore, J., *Curr. Opin. Biotechnol.*, 10 (1999) 54.
- Roberts, G., *Curr. Opin. Biotechnol.*, 10 (1999) 42.
- Mao, Q., Schunk, T., Gerber, B. and Erni, B., *J. Biol. Chem.*, 270 (1995) 18295.
- Furste, J.P., Pansegrau, W., Frank, R., Blocker, H., Scholz, P., Bagdasarian, M. and Lanka, E., *Gene*, 48 (1986) 119.
- Sambrook, J., Fritsch, E.F. and Maniatis, T., *Molecular Cloning: A laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989).
- Curtis, S.J. and Epstein, W.J., *Bacteriol.*, 122 (1975).
- Stüber, D., Matile, H., and Gorotta, G., in I. Lefkovits, and Pernis, B. (Eds.), *System for high level production in Escherichia coli and rapid purification of Recombinant Proteins: Application to Epitope mapping, Preparation of Antibodies, and Structure-Function Analysis*. Immunological Methods pp. 121-152.
- Bartels, C., Xia, T.-h., Billeter, M., Güntert, P. and Wüthrich, K.J., *Biomol. NMR*, 6 (1995) 1.
- Wüthrich, K. *NMR of Proteins and Nucleic Acids*, 1st ed., Wiley, New York, NY (1986).
- Güntert, P., Mumenthaler, C. and Wüthrich, K.J., *Mol. Biol.*, 273 (1997) 283.
- Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham III, T.E., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Ferguson, D.M., Radmer, R.J., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A., *AMBER 5.0* (1997) University of California, San Francisco.
- Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A., *J. Am. Chem. Soc.*, 117 (1995) 5179.
- Herzberg, O., *J. Biol. Chem.*, 267 (1992) 24819.
- Frisch, M.J., Trucks, G.W., Schlegel, H.B., Gill, P.M.W., Johnson, B.G., Robb, M.A., Cheeseman, J.R., Keith, T.A., Petersson, G.A., Montgomery, J.A., Raghavachari, K., Al-Laham, M.A., Zakrzewski, V.G., Ortiz, J.V., Foresman, J.B., Peng, C.Y., Ayala, P.A., Wong, M.W., Andres, J.L., Replogle,

- E.S., Gomperts, R., Martin, R.L., Fox, D.J., Binkley, J.S., Defrees, D.J., Baker, J., Stewart, J.P., Head-Gordon, M., Gonzalez, C. and Pople, J.A., Gaussian 94 Revision C.3, (1995) Gaussian Inc., Pittsburgh PA
43. Bayly, C.I., Cieplak, P., Cornell, W.D. and Kollman, P.A., J. Phys. Chem., 97 (1993) 10269.
44. Cieplak, P., Cornell, W.D., Bayly, C.I. and Kollman, P.A., J. Comput. Chem., 16 (1995) 1357.
45. Mehler, E.L. and Solmajer, T., Protein Eng., 4 (1991) 903.
46. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E.J., Phys. Chem., 21 (1953) 1087.
47. Kraulis, P.J., Applied. Crystallogr., 24 (1991) 946-950.
48. Koradi, R., Billeter, M. and Wuthrich, K., J. Mol. Graphics, 14 (1996) 51.
49. Islam, S.A. and Sternberg, M.J.E., PREPI v.0.95 (1998) Biomolecular modelling Laboratory, Imperial Cancer Research Fund, London, UK (<http://www.icnet.uk/bmm>)
50. <http://www.povray.org>