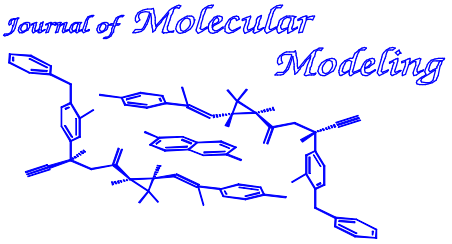
J. Mol. Model. **1995**, *1*, 150 – 160



Comparison of a Homology Built Model of Angiogenin to its Crystal Structure

Andrew D. Allen, Brendan J. Howlin * , Graham A. Webb,

Department of Chemistry, University of Surrey, Guildford, Surrey, GU2 5XH, UK (chp1aa@surrey.ac.uk)

Received: 30 June 1995 / Accepted: 3 August 1995

Abstract

The comparison of our homology built model of human angiogenin with the recently determined x-ray structure of the same is reported. The basic details of the structure in terms of alpha-helices and beta sheets were found to be common. The main differences between the model and the x-ray data lie in a c-terminal rearrangement in the x-ray structure that causes the c-terminus to end in a 3_{10} helix which puts the residue GLN-117 (ALA-122 in bovine pancreatic ribonuclease A, RNaseA) into the active site. The homology model was updated by producing a new sequence alignment using the information from the x-ray data which improved the r.m.s. by 0.5Å. This new alignment is also reported here. A check for systematic bias was carried out using the RNaseA structures from which the x-ray and homology models were derived. A detailed comparison of torsion angles and hydrogen bonding between all the structures have been compared and the model displays several hydrogen bonds that are not present in the parent RNaseA structures but are present in the x-ray structure of angiogenin.

Keywords: Angiogenin, Homology, Ribonuclease A

Abbrevations: *RNaseA*: crystal structure of bovine pancreatic ribonuclease A, *1ang*: crystal structure of angiogenin, *3rn3*: *RNaseA* used to produce our homology model, *5rsa*: *RNaseA* used to solve *1ang*, *angm*: Our homology built structure of angiogenin, r.m.s.: root mean square

Introduction

Angiogenin is a basic single chain protein of 123 amino acid residues, which is implicated in the growth of blood vessels in humans (Angiogenesis). The growth of blood vessels is important in diseases as varied as diabetic retinopathy and the growth of solid tumours. Angiogenin was first isolated from human colorectal adenocarcinoma line HT-29 [1]. Angiogenin is a member of the ribonuclease super-family and its ribonucleolytic activity is implicated in its angiogenic activity. It bears a close homology to bovine pancreatic ribonuclease A, having a percentage homology of 33% [2].

The crystal structure of bovine pancreatic ribonuclease A is known [3] along with crystal structures of the protein with inhibitors bound into the active site (see for example, Aguilar et al, 1992; Birdsall et al, 1992, Mills et al, 1992) [4,5,6] Hence a model of angiogenin based on the known structure of bovine pancreatic ribonuclease A would provide the basis for pharmacophore design for possible anti-cancer drugs. We have previously built such a model [7] and another model has also been proposed by Palmer [8]. The publication of the crystal structure of human angiogenin [9] offers the opportunity of comparing our model with this. Naturally, the crystal structure may not reflect the structure present in aqueous solution

* *To whom correspondence should be addressed*

due to crystal packing effects, etc., but will give some insight into the stability of structural elements present.

The purpose of this paper is to describe the comparison of our model (termed *angm*) with the crystal structure of human angiogenin (Brookhaven [10,11] protein database code *1ang*). In order to check that systematic bias is not occurring we also report the comparison with the crystal structures used for constructing our model, *3rn3* [12] and that used for molecular replacement to solve the human angiogenin structure (*5rsa*) [13]. An updated model for human angiogenin is constructed from this comparison and the changes are detailed. A recent comparison of the model produced by Scheraga's group [14] with the x-ray data has been published. This model used the ECEPP program which holds bond lengths and angles fixed, allowing variation in torsion angles only. Our model used the CHARMm forcefield where all parameters (including bond lengths and angles) were allowed to vary. The ECEPP formalism includes non-bonded interactions in terms of Lennard-Jones 12-6 potentials, hydrogen bonds are described by a 12-10 potential, and includes torsion potentials about peptide side-chains, and end-group bonds. Bond lengths and angles are fixed at standard values, characteristic of each type of amino acid residue. The forcefield has been described in the literature in detail [15]. The previous comparison showed that the essential features of the model were present in the x-ray data and that one helix was rotated by about 20º in the model. It was one of our aims in this comparison to see if allowing all bonds, angles and torsions to vary would cause any significant differences in our model. The ECEPP argument says that if bond lengths and angles are allowed to vary then energy minimisation can be achieved by varying a bond angle. When a bond angle is varied the resulting energy change is difficult for molecular mechanics energy minimisation to remove. This energy could more easily be taken out by varying torsion angles. Hence allowing bond angles and lengths to vary may result in incorrect conformations. The alternative argument is that if bond lengths and angles are fixed then the only way to relieve repulsive non-bonded interactions is by change of torsion angles, which can give rise to an artificial molecular conformation [16].

Methodology

Preparation of the proteins

All the data for the comparative study were collected on a Silicon Graphics Iris Indigo R4000 XZ running Insight II [17]. Although no energy or strain based calculations were carried out, the AMBER forcefield [18,19], as implemented, was used to determine the validity of each structure.

Porting the crystal structures

In order to compare the four proteins, *1ang*, *3rn3*, *5rsa* and *angm* (Figures 1, 2, 3 and 4) some preparative work needed to be done to enable superposition of one protein onto another. These preparations involve firstly the addition of hydrogens to the crystal structures where little or no data on the positions of hydrogens were available; secondly the charge state of basic and acidic residues needed to be standardised; thirdly the atom orders had to be the same. Whilst the resolution of the crystal structures used were insufficient to locate hydrogen positions, in order to utilise the software available to us hydrogens needed to be added. The errors associated with this procedure are only evident in the all atom r.m.s deviation comparisons, as hydrogens were ignored for the heavy atom and backbone comparisons. The homology model of angiogenin was generated using Quanta 3.0 by Polygen and CHARMm 21.3 [20] so hydrogens were already specified in this structure. For the other three proteins the addition of hydrogens was done automatically by a capping process [21]. This method adds hydrogens to unfilled valences as determined by existing bonds and geometry. This worked well for the majority of residue types, but some needed modification to their protonation states. The residue types that needed to be modified were the basic and acidic residues LYS, ARG, ASP and GLU. The HIS residues were left neutral with a hydrogen on the delta nitrogen, in preference to the epsilon nitrogen. All the basic and acidic residues were set to their ionisation states at neutral pH, therefore all LYS and ARG residues were protonated and all ASP and GLU residues deprotonated. The exception to this being the catalytic histidines, HIS 119 (114 in angiogenin) was protonated at both the delta and epsilon nitrogens, and HIS 12 (13 in angiogenin) was neutral with a proton at the delta nitrogen. This was consistent with the accepted mechanism of action of ribonuclease A [22]. The atom types and typing rules from the AMBER forcefield were used to check each protein. This located any unknown and therefore erroneous residues. In order to fully satisfy these typing rules lone pairs were added

Figure 1 (PDB-file: figure1.pdb): *Crystal structure of angiogenin, taken from* 1ang, *with hydrogens added and residues set to protonation states seen at pH 6.8*

Figure 2 (PDB-file: figure2.pdb): *Crystal structure of ribonuclease A, taken from* 3rn3, *with hydrogens added and residues set to protonation states seen at pH 6.8*

Figure 3 (PDB-file: figure3.pdb): *Crystal structure of ribonuclease A, taken from* 5rsa, *with unresolved hydrogens added and residues set to protonation states seen at pH 6.8*

Figure 4 (PDB-file: figure4.pdb): *Crystal structure of ribonuclease A, taken from our homology model* angm, *with residues set to protonation states seen at pH 6.8*

Figure 5 (PDB-file: figure5.pdb): *PDB file of deletions made at in* RNaseA *at SER-59 to ASN-71, GLU-58 to ASN-68 in angiogenin, needed to align the backbone sequences.*

Table 1 *R.M.S. comparison between Angiogenins (native versus model)*

Mode of comparison	R.M.S. deviation (Å)	Number of atoms
All atom	4.56	1970
Heavy atom	4.21	993
Backbone	3.14	492

Table 2 *R.M.S. comparison between Ribonuclease A structures (*3rn3 *versus* 5rsa*)*

Mode of comparison	R.M.S. deviation (Å)	Number of atoms
All atom	1.21	1872
Heavy atom	0.75	951
Backbone	0.18	496

to all CYS residues, these lone pairs were included in the R.m.s. comparisons as they provide information on the orientation of the disulphide bonds.

During the addition of hydrogens and modification of the basic and acidic residues the order of some of the atoms in the structure files had been altered for the modified residues. Owing to the different connectivity libraries used in the x-ray refinement programs and the AMBER package, the final modification made to the protein's data files was to reorder the atoms for each residue type, into a standard order throughout all four proteins. For some residues the naming of the hydrogens was erroneous, due to inconsistencies in crystallographic nomenclature and molecular mechanics atom order. Therefore, these atoms, particularly on isoleucine residues, needed modification, this also meant changing the atom naming in the connectivity tables. Once this was done the two *RNaseA* structure files were identical save for name and co-ordinates, as were the two angiogenin co-ordinate files. At this stage preliminary comparison of the two pairs of proteins was made (Tables 1 and 2) to check the atom reordering, and to ensure the two pairs of proteins were superimposable. Since the pairs of proteins, (*1ang angm*) and (*5rsa 3rn3*), could be directly overlaid atom by atom with no errors the protein files had been correctly prepared.

Sequence alignment

The most important task in preparing the proteins to be analysed, was deciding how to best align the sequences of the two pairs of different proteins, for *RNaseA* and angiogenin. A poor sequence alignment would over estimate the differences between the proteins compared. With the crystal structures of angiogenin and *RNaseA*, it was possible to overlay the backbone of the two proteins. By looking at selected portions of the two chains and overlaying them, it was possible to arrive at a much more accurate sequence alignment (Table 3) than by simply considering the protein sequences.

The first few residues of the two chains are badly defined and the first match in the sequences comes at THR-3 in *RNaseA* which is structurally equivalent to SER-4 in angiogenin. This similarity continues until LYS-17 of *RNaseA* and SER-16 of angiogenin, where for five or six residues the backbone conformations differ significantly. This region is a loop in both structures. In *RNaseA* this loop extends for six residues, whereas in angiogenin it lasts for only five, such that the sequences match again at SER-22 in *RNaseA* to ASP-22 in angiogenin. To align the backbones for comparison ALA-20 of *RNaseA* was removed allowing SER-21 in *RNaseA* to match ARG-21 in angiogenin. The next difference is at LYS-37 to CYS-40 in *RNaseA*, SER-37 to CYS-39 in angiogenin. Here a 4-turn occurs in *RNaseA* which takes four residues to complete, whereas in angiogenin there is a 3-turn lasting three residues. After this the chains realign with CYS-40 of *RNaseA* matching CYS-39 angiogenin. Thus the deletion of ASP-38 in *RNaseA* has been made to realign the sequences. The next major difference in the two proteins occurs in the region *RNaseA* SER-59 to ASN-71, angiogenin GLU-58 to ASN-68 (refer to Figure 5). Firstly there is a 4-turn in the angiogenin backbone covering four residues where there is no equivalent in *RNaseA*; then the *RNaseA* chain has an extended section of beta sheet incorporating a 3-turn, taking five more residues than angiogenin. The residues in between these two differences, GLN-60 to ALA-64 in *RNaseA* and GLY-62 to ARG-66 in angiogenin, can be aligned, as can the residues either side of both these differences. To compare the sequence in this region two deletions are made, ASN-59 to ASN-61 were removed from angiogenin, and CYS-65 to GLN-69 in *RNaseA*. Another minor discrepancy occurs from residues THR-87 to PRO-93 in *RNaseA*, HIS-84 to PRO-90 in angiogenin where angiogenin has a loop region, and *RNaseA* has a beta sheet and beta turn. Since the loop covers the same number of residues as the beta turn and sheet, i.e. seven residues, no deletions need to be made. From this point onwards the chains continue to align up, until GLY-112 to VAL-116 in *RNaseA*, ASN-109 to LEU-111 in angiogenin. Here the last major difference is in a region of beta sheet that extends for two extra residues in *RNaseA*, including a 5-turn, whereas angiogenin has a 3-turn. To realign the sequences ASN-113 and PRO-114 were removed from *RNaseA*. The rest of the sequences match until ALA-122 of *RNaseA* and GLN-117 of angiogenin, which forms the N-terminus and extends into the solvent, and is therefore poorly resolved.

RNaseA		K	E	T	A	A	A	K	F	E	R	Q	H	M	D	S	S	T	S	A	A	S	S	S
											10										20			
angm	Q	D	N	S	R	Y	T	H	F	L	T	Q	H	Y	D	A	K	P	Q	G	R	D	D	R
										10										20				
1ang	Q	D	N	S	R	Y	T	H	F	L	T	Q	H	Y	D	A	K	P	Q		G	R	D	D
										10											20			

RNaseA	N	Y	C	N	Q	M	M	K	S	R	N	L	T	K	D	R	C	K	P	V	N	T	F	V
							30										40							
angm		Y	C	E	S	I	M	R	R	R	G	L	T	S	P		C	K	D	I	N	T	F	I
							30											40						
1ang	R	Y	C	E	S	I	M	R	R	R	G	L	T	S		P	C	K	D	I	N	T	F	I
							30											40						

RNaseA	H	E	S	L	A	D	V	Q	A	V	C	S				Q	K	N	V	A	C	K	N	G
																60								
angm	H	G	N	K	R	S	I	K	A	I	C	E				N	K	N	G	N	P	H	R	E
				50													60							
1ang	H	G	N	K	R	S	I	K	A	I	C	E	N	K	N	G	N	P	H	R				
				50										60										

RNaseA	Q	T	N	C	Y	Q	S	Y	S	T	M	S	I	T	D	C	R	E	T	G	S	S	K	Y
		70										80										90		
angm			N	L	R	I	S	K	S	S	F	Q	V	T	T	C	K	L	H	G	G	S	P	W
					70										80									
1ang		E	N	L	R	I	S	K	S	S	F	Q	V	T	T	C	K	L	H	G	G	S	P	W
					70										80									

RNaseA	P	N	C	A	Y	K	T	T	Q	A	N	K	H	I	I	V	A	C	E	G	N	P	Y	V
								100										110						
angm	P	P	C	Q	Y	R	A	T	A	G	F	R	N	V	V	V	A	C	E	N	G			L
	90										100										110			
1ang	P	P	C	Q	Y	R	A	T	A	G	F	R	N	V	V	V	A	C	E	N			G	L
	90										100												110	

RNaseA	P	V	H	F	D	A	S	V																
				120																				
angm	P	V	H	L	D	Q	S	I	F	R	R	P												
									120															
1ang	P	V	H	L	D	Q	S	I	F	R	R	P												
									120															

Table 3: *Sequence alignment for* RNaseA *against the sequence alignment used to build the homology model of angiogenin and the sequence alignment of the angiogenin crystal structure to* RNaseA *from a three dimensional point of view*

Conserved atom set

In order to compare a *RNaseA* structure to an angiogenin structure an atom set of conserved atoms was generated for each protein. This was done for two cases; a) the original sequence alignment used to generate the homology model *angm*; b) the alignment detailed above based on a 3D alignment. Firstly the conserved residues from the sequence alignment were added. Secondly the backbone atoms and conserved sidechain atoms of non-conserved residues were added noting the extent of sidechain similarity. For example LYS compared to ASP are similar up to the gamma carbon, after this point ASP has two oxygens whereas LYS continues with a methylene. Here the beta carbon and hydrogens plus the gamma carbon were included in the comparison. For all four proteins a conserved atom list of 1297 atoms, (1346 for the original alignment), including hydrogens was generated. Angiogenin itself has 1970 atoms and this represents a similarity of 65.8% on an atom by atom comparison.

The remainder of the data collected in this study uses the subsets of conserved atoms generated for each protein. Data was collected on a residue by residue basis and each was compared to the other three proteins to give six sets of data, relating to r.m.s. deviation after superposition and backbone psi and phi torsion angles. A set of hydrogen bonds was also generated for each protein, to compare conserved secondary structural features. For the r.m.s. deviations and torsion angle data a macro was written in Biosym's command language to automate data collection. The hydrogen bonds were measured with an angle cut off of 120°, and distance cut offs of 2.5Å between donor hydrogens and nitrogen, oxygen and sulphur acceptors.

Table 4: *Conserved atom superpositions, R.M.S. deviations in Å using the sequence alignment determined for the homology model, and by 3D structure comparison*

		Sequence alignment						**difference**		
		Homology model			**3D structure**			**Homology – 3D**		
		1ang	**5rsa**	**3rn3**	**1ang**	**5rsa**	**3rn3**	**1ang**	**5rsa**	**3rn3**
angm	All	3.55	1.65	1.52	3.07	2.27	2.20	0.48	-0.62	-0.67
	Heavy	3.25	1.41	1.36	2.89	2.13	2.11	0.36	-0.72	-0.75
	Backbone	2.93	1.29	1.30	2.77	2.11	2.12	0.16	-0.82	-0.82
	all-heavy	0.30	0.23	0.16	0.18	0.14	0.08			
3rn3	All	3.52	0.83		2.47	0.84		1.05	-0.01	
	Heavy	3.24	0.47		2.19	0.47		1.05	-0.01	
	Backbone	2.93	0.18		2.03	0.18		0.90	0.00	
	all-heavy	0.28	0.36		0.28	0.36				
5rsa	All	3.51			2.44			1.07		
	Heavy	3.25			2.18			1.06		
	Backbone	2.93			2.02			0.91		
	all-heavy	0.27			0.26					

Results

The comparisons of the proteins as a whole and by residue has been carried out using the two sequence alignments for angiogenin to *RNaseA*. The first is the sequence alignment used to build *angm*, which was based on the amino acid sequences and results of mutagenesis studies. The second is one based on the three dimensional overlay of backbone atoms for the two proteins. Each comparison has been carried out using a subset of atoms for each protein which corresponds to the conserved atoms between angiogenin and *RNaseA* for these sequence alignments.

R.M.S. deviations after superposition

To check if the hydrogens added to the crystal structures significantly influenced the overall r.m.s. deviation, the superposition's have been carried out with an all atom and a heavy atom mode where hydrogens are ignored. To check for tertiary structure agreement a backbone comparison was used where all sidechains are neglected.

Protein vs. Protein comparison

The addition of hydrogens to the crystal structures *1ang*, *3rn3* and the missing hydrogen data in *5rsa*, increases the r.m.s. deviations in the comparisons (Table 4) by the range 0.08 to 0.36Å, this indicates that the addition of hydrogens to the models causes the structures to deviate more. This increase is only significant in the comparison of the two *RNaseA* structures, where the increased deviation represents 43% of the entire r.m.s. deviation. This deviation is due to a difference in hydrogen models, *5rsa* has some hydrogens determined by neutron diffraction, whereas all hydrogens for the *3rn3* structure were placed in calculated positions. The increase in deviation for other comparisons was found to be insignificant considering about 600 hydrogens have been added to the overall structure.

The two sequence alignments show two trends. The first trend appears in both sequence alignments used, and shows that *angm* is more similar to *RNaseA* than to *1ang*. However the second trend seen in the sequence alignment based on the 3D backbone overlay of the two proteins, shows that there is an increase in the deviation between *angm* and *RNaseA*, and a decrease in the deviation of *angm* to *1ang*. This indicates that although *angm* was built by homology to

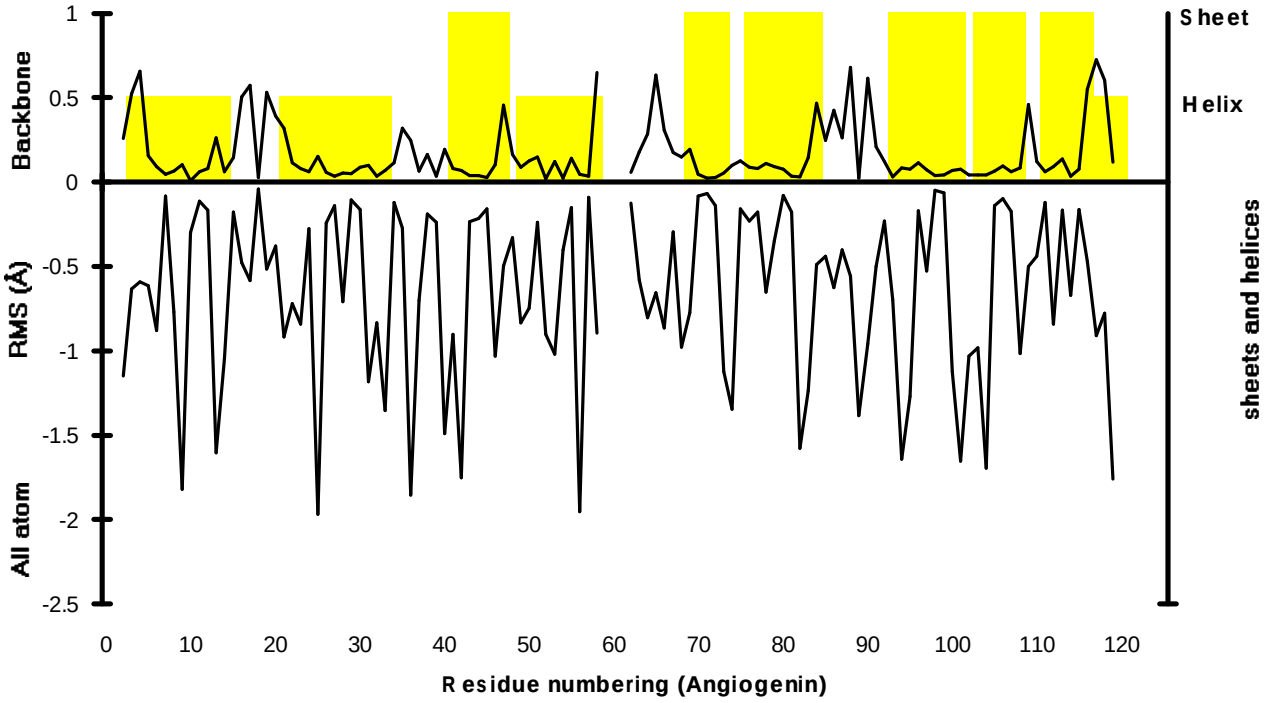


Figure 6: *Graph of r.m.s. deviation in Å of the co-ordinates by superposition of* 1ang *and* angm, *conserved atoms only, backbone and all atom representation.*

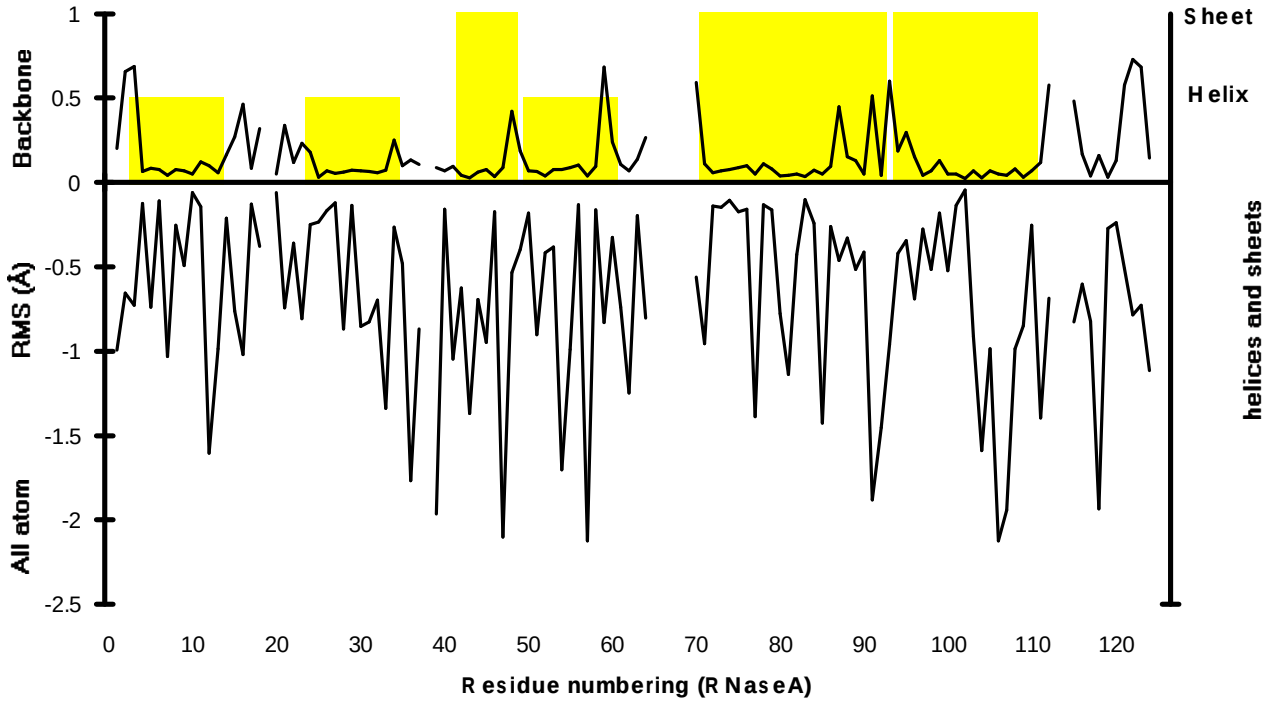


Figure 7: *Graph of r.m.s. deviation in Å of the co-ordinates by superposition of* 5rsa *and* 1ang, *conserved atoms only, backbone and all atom representation.*

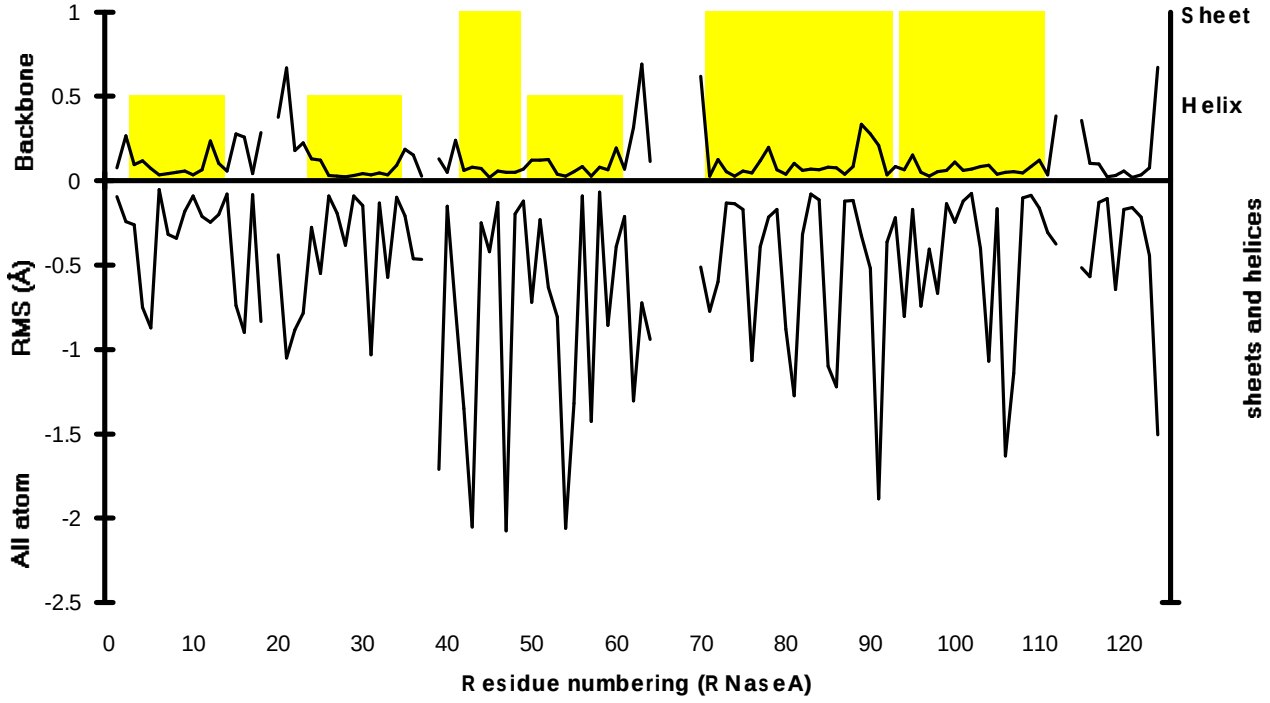


Figure 8: *Graph of r.m.s. deviation in Å of the coordinates by superposition of* 3rn3 *and* angm, *conserved atoms only, backbone and all atom representation.*

RNaseA a more accurate sequence alignment reveals an increase in similarity to *1ang*.

Analysis of the heavy atom and all atom superpositions, indicates that the major differences in the proteins comes from the tertiary or backbone structure and not sidechain conformations. Although sidechains do cause deviations the majority of sidechains are not in conserved positions, and therefore cause a uniform deviation. The notable exception to this is the comparison of the two *RNaseA* structures *5rsa* and *3rn3* where the backbones are almost identical, and the only major deviations occur in sidechain conformations. Table 1 shows that the two *RNaseA* structures are the most similar, as expected. An all atom comparison of the conserved atom set gives an r.m.s. deviation that is less than 1Å. The structure of angiogenin built by homology to *RNaseA* is more similar to the *RNaseA* structure that was used to generate it, and has a higher deviation from the reported crystal structure of angiogenin (*1ang*). The crystal structure of angiogenin itself is markedly different from all the other three protein structures, indicating that despite a high homology to *RNaseA*, main chain and sidechain positions on a global scale are distinctly different. For a more in depth comparison, the differences have also been calculated on a residue by residue basis.

Residue by residue comparison

Several tertiary features of *RNaseA* are slightly different in the corresponding angiogenin sequence. Some of these features have been correctly modelled in *angm* but not others. There are three helical regions in *RNaseA*, and these occur in almost identical positions in angiogenin. For these regions the main deviation in the backbone sequence occurs at the turns into and out of the helices. Similarly the beta sheet regions are highly conserved though some distinct differences do appear. Refer to figures 6, 7, and 8 to follow the following differences:

Between SER-15 and ASN-24 of *RNaseA*, ALA-16 to ARG-24 in angiogenin, there is significant backbone deviation. This is due to a flexible loop region, connecting the first two helices.

Between ASN-34 and CYS-40, GLY-34 to CYS-39 in angiogenin, following the second helix there is another loop region which leads into a strand of beta sheet incorporating a 4-turn in *RNaseA* and a 3-turn in angiogenin, held into place by a disulphide bond. This region is flexible enough to cause significant differences between angiogenin and *RNaseA*.

At GLU-49 in *RNaseA*, GLY-48 in angiogenin, there is a change from a sheet into a helix. In this short region GLY-48, angiogenin, is twisted with respect to GLU-49, *RNaseA*, causing a significant r.m.s deviation.

Between SER-59 and GLN-60 in *RNaseA*, GLU-58 and ASN-61 in angiogenin, there is a 4-turn in angiogenin which is not present in *RNaseA*.

Between residues ALA-64 and THR-70 in *RNaseA*, ARG-66 and GLU-67 in angiogenin, *RNaseA* has an extended beta sheet with a 3-turn coming back to match up with the angiogenin backbone at residue GLU-67 in angiogenin, THR-70 in *RNaseA*.

Between GLU-86 and ASN-94 of *RNaseA*, LEU-83 to PRO-91 in angiogenin, there is another flexible loop region in angiogenin, which corresponds to a beta turn and beta sheet in *RNaseA*, and this gives rise to significant r.m.s. deviation.

Between CYS-110 and VAL 116 of *RNaseA*, CYS-107 to LEU-111 in angiogenin, *RNaseA* has a more extended beta sheet ending in a 5-turn, whereas in angiogenin there is a 3-turn.

Table 5: *Active site comparison*

Angiogen residues	**13, 40, 42, 43, 44, 114, 115, 116, 118**		
RNaseA residues	**12, 41, 43, 44, 45, 119, 120, 121, 123**		
		1ang	***5rsa***	***3rn3***
angm	atom	2.52	1.29	1.08
	heavy	2.16	0.82	0.78
	backbone	1.83	0.50	0.51
3rn3	atom	2.33	0.89	
	heavy	2.02	0.43	
	backbone	1.75	0.13	
5rsa	atom	2.25		
	heavy	1.97		
	backbone	1.72		

Comparison of active site by r.m.s

The r.m.s. values are lower for the active site residues alone, thus indicating a higher degree of similarity within the active site for the proteins compared. The r.m.s. values of superimposed active site residues, (refer to Table 5), shows a similar trend to the comparison of the proteins as a whole. *Angm*'s active site being more similar to the active site conformation of both *RNaseA* structures, and *1ang* being markedly different to both our model (*angm*) and to *RNaseA*. However when *angm* was constructed the C-terminal conformation was unknown. This combined with unexpected sidechain conformations on residues such as GLU-117 in angiogenin, causes the majority of the differences seen in our model.

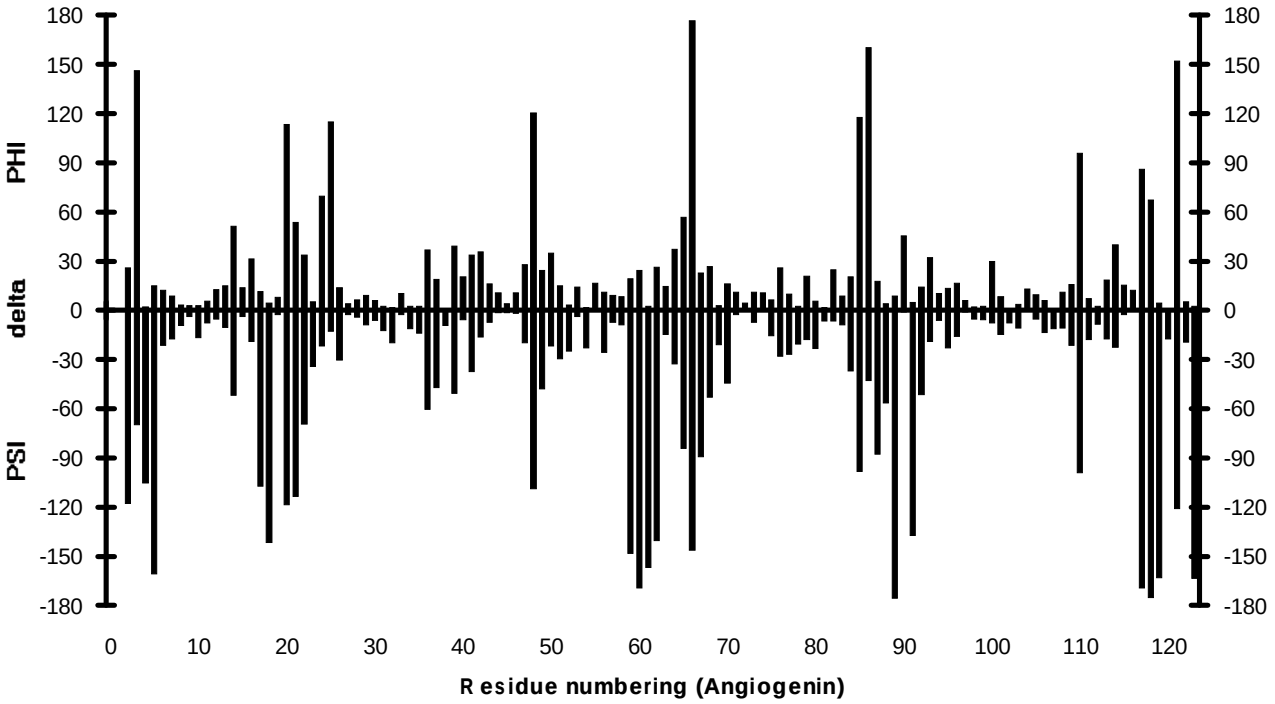


Figure 9: *Graph of backbone torsion (psi and phi) angle differences for* angm *and* 1ang.

Torsion Angles

Comparing the phi and psi backbone torsion angles for the proteins produces similar patterns to those seen in the r.m.s. deviation data (Figure 9). As expected the comparison of *3rn3* to *5rsa* yields a very close match in torsion data, the highest differences in the phi and psi angles being less than 30º, with an average difference less than 5º. The comparison for the angiogenin and *RNaseA* torsion angles yields differences that range up to 180º. These large differences occur in flexible loop regions, and in places where the homology of angiogenin to *RNaseA* breaks down. In the region 64-70 in *RNaseA* the sequences cannot be aligned so this region is left blank in the figure.

Conserved regions in the backbone structure are easily picked out as areas of low torsion angle difference between the proteins compared. As with the above comparison on r.m.s. deviation these regions occur where helices and beta sheet are matched between angiogenin and *RNaseA.*

Intramolecular Hydrogen bonds

The hydrogen bonding data helps to pick out regions where the secondary structure of the proteins under study differ. If a sheet region is extended or a helix is a different length, then the hydrogen bonds should indicate this. The loop regions will be indicated as areas of no intramolecular hydrogen bonding, a sheet region will connect strands of protein that run parallel to each other, and a helix will have sections of backbone that form hydrogen bonds to residues that are four units back. All these features can be picked out and compared.

Backbone hydrogen bonding

Backbone hydrogen bonds are highly conserved within helical and sheet regions of angiogenin and *RNaseA.* For the first helix THR-3 to MET-13 of *RNaseA*, SER-3 to TYR-14 in angiogenin, the hydrogen bonds of the two *RNaseA* backbones are identical, as expected. *1ang* has hydrogen bonds in all equivalent positions and also has three extra ones. One of these extra hydrogen bonds is conserved in our model (*angm*), and represents an extended helix from the start of the chain from residues GLN-1 and ASP-2. The second helix SER-21 to ARG-33 of *RNaseA*, ASP-22 to ARG-33 of angiogenin, again shows a high degree of conservation in hydrogen bond positions. Of those seen in the *RNaseA* structures only one hydrogen bond is absent on comparison between the two *RNaseA* structures and this occurs in *3rn3*. *1ang* matches hydrogen bonds in all equivalent positions in the *RNaseA* backbone in this region, and *angm* conserves all but one. *Angm* also makes three interhelical hydrogen bonds which do not appear in any of the

other three proteins. The last conserved helix, SER-50 to GLU-60 in *RNaseA* and ASN-49 to GLU-58 in angiogenin, is very highly conserved, the only minor differences being at the start of the helix.

The sheets are also highly conserved with 38 hydrogen bonds being made and conserved in both *RNaseA* structures, with only 3 hydrogen bonds not being matched. *1ang* matches to 29 of these 38 hydrogen bonds and *angm* matches to 28. The differences seen between *1ang* and *RNaseA* occur at three separate places. The first of these ALA-64 to ASN-71 in *RNaseA*, ARG-66 to ASN-68 angiogenin, corresponds to the putative cell binding site of angiogenin [23]. The second GLY-88 to ALA-96, GLY-85 to GLN-93 angiogenin, corresponds to a flexible loop in angiogenin but a turn and beta sheet in*RNaseA*. The last difference is at residues ALA-122 to VAL-124, GLN-117 to ILE-119 angiogenin, here the C-terminus of *1ang* is poorly defined and is therefore expected to be different. The model structure *angm* has similar differences to the *RNaseA* structures, the first at GLN-60 THR-70 *RNaseA*, GLY-62 to GLU-67 angiogenin, also corresponds to the cell binding site though this difference covers more residues. The second region of difference for *angm* falls in an almost identical position to one seen in *1ang* at GLY-88 to PRO-93, GLY-85 to PRO-90 angiogenin, and also corresponds to a loop region in *angm*. The last difference for *angm* is from GLU-111 to VAL-116, GLU-108 to LEU-111 in angiogenin, and would appear to coincide with another portion of angiogenin thought to be involved in cell binding. However *1ang* shows no appreciable difference in hydrogen bonds made in this region compared to *RNaseA*.

Active site hydrogen bonds

There are differences in the hydrogen bonding to the active site residues in *RNaseA* and Angiogenin. This is as a consequence of the C-terminal rearrangement noted in the angiogenin crystal structure. As there was no information about this rearrangement when our model was built, our model therefore more closely resembles the *RNaseA* case than that seen in the angiogenin crystal structure, *1ang*. As noted, in *RNaseA* the catalytically important LYS-41 makes no hydrogen bonds or salt bridges in either our model or the angiogenin crystal structure. Interestingly our model correctly predicts the hydrogen bonding supporting network in the vicinity of THR-45 although hydrogen bonds to the sidechains from the C-terminus to this region are not conserved between our model and the angiogenin crystal structure. One interaction that is present in our model and the angiogenin crystal structure but not in the corresponding *RNaseA* structures is a backbonding hydrogen bond from the side-chain of GLU-67 to the backbone of residue NH of GLU-67. This residue's equivalent at this position in *RNaseA* is a THR .

The proposed cell binding site in Angiogenin comes close to the expected purine binding site, and has its hydrogen bonding correctly predicted by our model as the hydrogen bonding to residues 111 and 112 is identical. In speculation there is a possibility that during cell binding this region is altered, causing the purine binding pocket to become more highly defined. Although there is no specific experimental data to prove this recent work by Furumichi et al [24] shows antiangiogenic activity by adenosine 3',5'-cyclic monophosphates.

Conclusion

The technique of protein homology modelling has been shown to be able to produce structures that reasonably well reproduce experimental structures. R.m.s. values of about 1 Å have been reported in the literature [25]. Our model of human angiogenin was built before the crystal structure was known and the cell binding site of angiogenin which was proposed agrees in most of its structural features. One major difference is in the observed C-terminal rearrangement in the crystal structure which could not be predicted by homology modelling. This rearrangement may serve to protect endothelial cells from catalytically active angiogenin, rearranging to more closely resemble our model and the other ribonucleases before its catalytic action, or undergoing some rearrangement upon translocation to the nucleus of the proliferating endothelial cells. One other possibility is that this C-terminal rearrangement may not occur in solution, being induced by the crystal packing environment seen in the x-ray data in *1ang*. Obviously solution NMR is needed to deal with these possibilities and some work to this end is ongoing on bovine angiogenin [26].

The overall r.m.s. of our model of 3.5.Å is affected by this rearrangement but it is interesting to note the degree to which the secondary structure is conserved and that our model is able to predict subtle effects of the hydrogen bonding which are absent in the parent *RNaseA*.

A comparison of our model (*angm*) with the homology model data reported by Allen et al. [14], shows that overall, the r.m.s. deviations of *angm* compared to the reported crystal structure (*1ang*) are smaller than those previously reported. The r.m.s. deviations for our model to the crystal structure are 3.55Å for all atoms and 2.93.Å for the main chain atoms only. The corresponding values for the previous model (here denoted as angs) are 4.44Å and 3.15Å respectively. It should be noted however that the angs model was constructed in 1986 using proline parameters that have been superseded. These parameters caused a large deviation between the model and the crystal structure in the region of residues 15-22. This was caused by incorrect assignment of dihedral angles for proline 18. The angs model does not report a helix in the region 50-56, helix 3 in *angm*, using the program DSSP. Although it is reported the residues are in an approximately helical conformation. Sheets 5 and 6 in *angm*, (residues 103-108 and 111-118), are closer to the crystal structure than in the angs model where the sheets are significantly shorter. Both models have missed the sheet in the crystal structure between residues 62-65. Interestingly the sequence align-

ments used for both homology models were identical. Both forcefield approaches have produced models that are reasonable representations of the crystal structure although unexpected deviations in the crystal structure led to similar errors in prediction. Overall the *angm* model more closely represents the crystal structure on the basis of r.m.s. deviation and secondary structure prediction. However it should be borne in mind that the angs structure is the product of 1986 technology and a newer model using more recent software may be much closer to the crystal structure.

The solving of the crystallographic structure of native angiogenin provides important insight into the mode of action of angiogenin, and may lead to an understanding of the mechanism of action. This opens up avenues of further work in this field towards producing effective inhibitors and other control drugs that will eventually lead to better treatment of disorders caused by uncontrolled Angiogenesis. A full understanding of the physiological action of angiogenin, with reference to it's structure, will greatly facilitate the design of novel inhibitors.

Acknowledgements Computational results obtained using software programs from Biosym Technologies of San Diego, and graphical displays were printed out from the *Insight®II* molecular modelling system. We would also like to thank the Association for International Cancer Research (AICR) for funding.

References

1. Fett, J.W.; Strydom, D J.; Lobb, R.R.; Alderman, E.M.; Bethume, J.L.; Riordan, J.F.; Vallee, B. L. *Biochemistry* **1985**, *24*, 5480.
2. Strydom, D.J.; Fett, J.W.; Lobb, R.R.; Alderman, E.M.; Bethune, J.L.; Riordan, J.F.; Vallee, B.L. *Biochemistry* **1985**, *24*, 5486.
3. Borkakoti, N.; Moss, D.S.; Palmer, R.A. *Acta Cryst B* **1982**, *38*, 2210
4. Aguilar, C.F.; Thomas, P.J.; Mills, A.; Moss, D.S.; Palmer, R.A. *J. Mol. Biol.* **1992**, *224*, 265.
5. Birdsall, D.L.; M^{c}Pherson, A. *J. Biol. Chem*. **1992**, *267*, 22230.
6. Mills, A.; Gupta, V.; Spink, N.; Lisgarten, J.; Palmer, R.A.; Wyns, L. *Acta. Cryst. B* **1992**, *48*, 549.
7. Chen, J.; Howlin, B.J.; Tomkinson, N.P.; Webb, G.A. *J. Chem. Cryst.* **1994**, *24*, 27.
8. Palmer, K.A.; Scheraga, H.A.; Riordan, J.F.; Vallee, B.L. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 1965.
9. Acharya, K. R.; Shapiro, R.; Allen, S.C.; Riordan, J.F.; Vallee, B.L. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2915.
10. Abola, E.E.; Bernstein, F.C.; Bryant, S.H.; Koetzle, T.F.; Weng, J. In Allen, F. H.; Bergerhoff, G.; Sievers, R. (Eds.) *Crystallographic databases - Information Content, Software Systems, Scientific Applications*, Data Commision of the International Union of Crystallography, Cambridge 1987, pp 107-132.
11. Bernstein, F.C.; Koetzle, T.F.; Williams, G.B.J.; Meyer, -Jr,E.F.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J Mol. Biol*. **1977**, *112*, 535.
12. Howlin, B.J.; Moss, D.S.; Harris, G.W. *Acta. Cryst. Sect. .A* **1989**, *45*, 851.
13. Wlodawer, A.; Sjolin, L. *Biochemistry* **1983**, *22*, 2720.
14. Allen, S.C.; Acharya, K.R.; Palmer, K.A.; Shapiro, R.; Vallee, B.L.; Scheraga, H.A. *J. Prot. Chem.* **1994**, *13*, 649.
15. Sippl, M.J.; Némethy, G.; Sheraga, H.A. *J. Phys. Chem*. **1984**, *88*, 6231.
16. White, D.N.J.; Ruddock, J.N.; Edgington, P.R.; Molecular Mechanics. *Computer-Adied Molecular Design* ed. Richards, W.G.; IBC Technical Services Ltd. London, 1989.
17. InsightII, vers 2.3.5, San Diego: Biosym Technologies, 1993.
18. Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S.Jr.; Weiner, P. *J. Am. Chem. Soc*. **1984**, *106*, 765.
19. Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. *J. Comp. Chem.* **1986**, *7*, 230.
20. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 8234.
21. Builder module. *Insight II User Guide*, version 2.3.0; Biosym Technologies: San Diego 1993..
22. Fersht, A. *Enzyme structure and mechanism 2nd Edn*.; W.H. Freeman and Co.: New York, 1985.
23. Moroianu, J.; Riordan, J.F. *Proc. Natl. Acad. Sci. USA* **1994**, *91:5*, 1677.
24. Furumichi, T.; Yamada, Y.; Suzuki, T.; Furui, H.; Yamauchi, K.; Saito, H. *Jpn. Heart J* **1992**,. *33*, 371.
25. Blundell, T.L.; Carney, D.; Gardner, S.; Hayes, F.; Howlin, B.; Hubbard, T.; Overington, J.; Singh, D.A.; Sibanda, B.L.; Sutcliffe, M. *Eur. J. Biochem* **1988**, *172*, 513.
26. Reisdorf, C.; Abergel, D.; Bontems, F.; Lallemand, J. Y.; Decottignies, J.P.; Spik, G. *Eur. J. Biochemistry* **1994**, *224:3*, 811.