Computer-Assisted Modeling and Examination of Protein-Molecule Complexes*

Klaus Müller
Central Research Units, F. Hoffmann-La Roche & Co. Ltd., CH-4002 Basel/Switzerland

Abstract

Three applications of the new Roche Interactive Molecular Graphics system are outlined. They include (i) the computer-assisted fitting of various brodimoprim derivatives into the active sites of L.casei and E.coli dihydrofolate reductases (DHFR), (ii) the modeling of complexes of L.casei DHFR with its natural substrate and reaction product, dihydrofolate and tetrahydrofolate, respectively, and (iii) the docking of the cofactor NADPH into the binary complex of E.coli DHFR and methotrexate.

Some features of the Roche Interactive Molecular Graphics (RIMG) system are discussed. This software package [1] is being developed at *Hoffmann-La Roche,* Basel, on a DEC VAX-11/780 computer connected to (i) an *Evans & Sutherland* Color Multi-Picture System with dials, function switches, a data tablet, and a VT100-type control terminal as parallel interactive devices and (ii) a Tektronix TEK-4113 color raster terminal with a TEK-4662/31 eight-color XY-plotter. RIMG provides various facilities for interactive examination of protein structures and cavity analyses, for modeling, superposition, and comparison of small molecules, and for design and evaluation of protein-molecule complexes. Efficient new algorithms for complete topological analyses of complex molecular structures enhance many features of molecular modeling and structure matching. An extended module for interactive and computer-assisted superposition of fully or partially flexible molecular structures with various optional constraints is a major asset of RIMG. Several graphic techniques are available for examination of intermolecular packing, such as dotted-surface, space-filling, chicken-wire, or body/surface-differentiated raster respresentations of molecules.

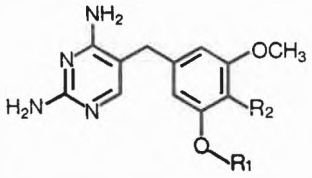
* Editorial remark: This article is a condensed version of the lecture that the author after having received the Werner-Award 1983 has given at the fall session of the Swiss Chemical Society on October 14, 1983 in Berne.

Three applications of the new graphics system are briefly discussed. They are based on refined X-ray structural data for complexes of *L.casei* and *E.coli* dihydrofolate reductases (DHFR) with methotrexate (MTX) and reduced nicotinamide adenine dinucleotide phosphate (NADPH), as obtained by the group of *J. Kraut* [2] and available through the Brookhaven Protein Data Bank.

All applications were performed at the Evans & Sutherland Color-MPS. A large set of color slides was taken directly from the monitor and used to document the discussions below. Since these slides cannot be reproduced here, a few relevant pictures were regenerated, using our Tektronix facilities, and are included here to illustrate the main points.

The first application involves a series of homologous brodimoprim derivatives [3,4], which exhibit a pronounced affinity to both *L.casei* and *E.coli* DHFR's (Fig. 1). Comparison of the binding constants shows that binding to L.casei DHFR does not change monotonically with increasing length of the alkanoic acid side chain, but reaches a maximum for the pentanoic acid derivative. By contrast, highest affinity for *E.coli* DHFR is obtained for the hexanoic acid derivative, and even the next higher homologue seems to be well accommodated by the enzyme. The biochemical data for brodimoprim derivatives and *E.coli* DHFR are well paralleled by those for corresponding trimethoprim derivatives [5].

In an attempt to rationalize the differences in chain length accommodation by the two enzymes, the five brodimoprim derivatives with two to six carbon atoms between the phenyl ring and the terminal caroxylate group were modeled, using X-ray structural data for trimethoprim [6] and standard structural parameters for the elongated side chains. These models



R_1	E. coli DHFR K_I (nM)		L. casei K_I (nM)
	R_2^*= OCH_3	R_2= Br	R_2= Br
CH_3	1.37	0.603	11.3
CO_2H	2.6	1.590	14.2
CO_2H	0.37	0.585	2.71
CO_2H	0.035	0.032	0.893
CO_2H	0.066	0.040	0.203
CO_2H	0.024	0.012	0.383
CO_2H	0.050	0.029	0.580

Fig. 1: Inhibitor constants for brodimoprim and brodimoprim derivatives [3,4] and corresponding trimethoprim derivatives [5].

were introduced into each DHFR by placing their diaminopyrimidine rings exactly into the positions, where methotrexate is known to bind [2].

Using multiparametric nonlinear optimization techniques, each brodimoprim molecule was then conformationally optimized in order to maximize the interaction of its terminal carboxylate group with Arg57 in either protein, while simultaneously minimizing conformational strain and interatomic collisions with the surrounding protein. During this process the protein was not allowed to relax and the diaminopyrimidine ring was held fixed in place for tight binding to Asp26 (*L.casei* DHFR) or Asp27 (*E.coli* DHFR).

The resulting conformations for the propionic, pentanoic, and heptanoic acid derivatives of brodimoprim are superimposed in Fig. 2 and 3 for the *L.casei* and *E.coli* cases, respectively. In both cases, the propionic acid side chain cannot reach Arg57. Similarly, the butyric acid side chain is still too short to form a symmetrical complex between the terminal carboxylate and the guanidinium group. Such a complex is easily achieved for the pentanoic acid derivative in either case. The higher homologues, however, are not easily accomodated within the *L.casei* DHFR. The longer side chains have to adopt rather puckered (strained) conformations in order to approach Arg57 properly

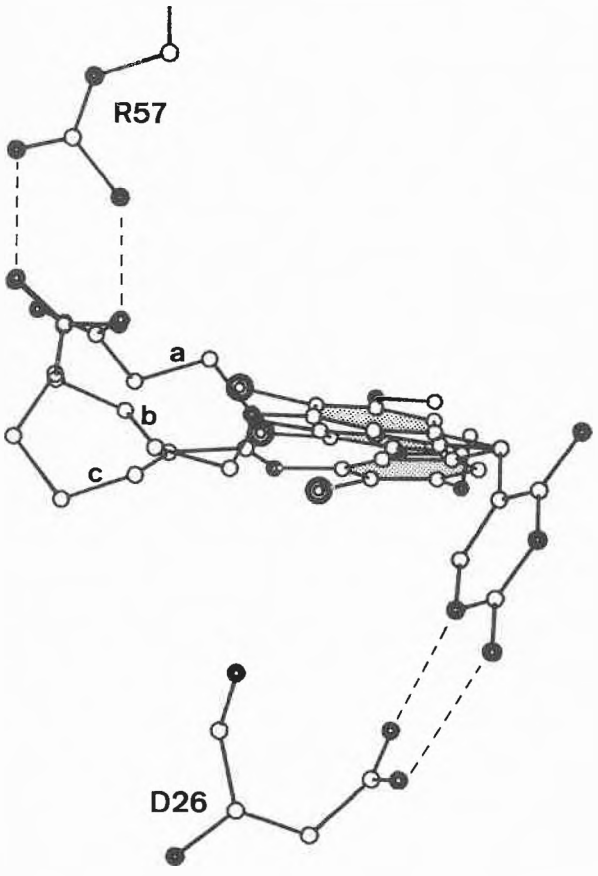


Fig. 2: Superposition of (a) propionic, (b) pentanoic, (c) heptanoic acid derivatives of brodimoprim, conformationally optimized to fit into the active site of *L.casei* DHFR; phenyl rings shaded.

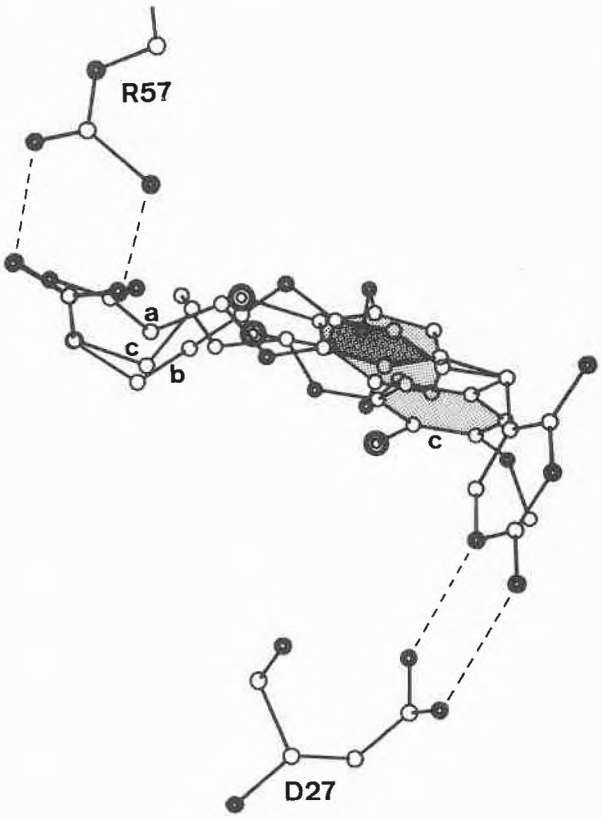


Fig. 3: Superposition of (a) propionic, (b) pentanoic, (c) heptanoic acid derivatives of brodimoprim, conformationally optimized to fit into the active site of *E.coli* DHFR; phenyl rings shaded.

within the spatial constraints of the enzyme cavity. This is particularly evident for the heptanoic acid side chain, which folds into a strained loop when forced to complex with Arg57.

By contrast, both the hexanoic and heptanoic acid derivatives are easily accommodated within the *E.coli* enzyme (Fig. 3) due to less stringent cavity constraints. Specifically, the tight cavity of *L.casei* DHFR confines the benzyl moiety to a relatively narrow conformational space, thus leaving little choice as to the optimum length for the side chain. *E.coli* DHFR, on the other hand, offers a wider cavity, allowing the benzyl moiety to swing back to accommodate a longer side chain.

While structural details of the computed conformations must remain hypothetical, qualitative aspects correlate well with available binding data. Our conclusions are further confirmed by recent X-ray crystallographic studies of binary complexes of structurally related propionic and hexanoic acid derivatives of trimethoprim with *E.coli* DHFR, using difference Fourier techniques [5].

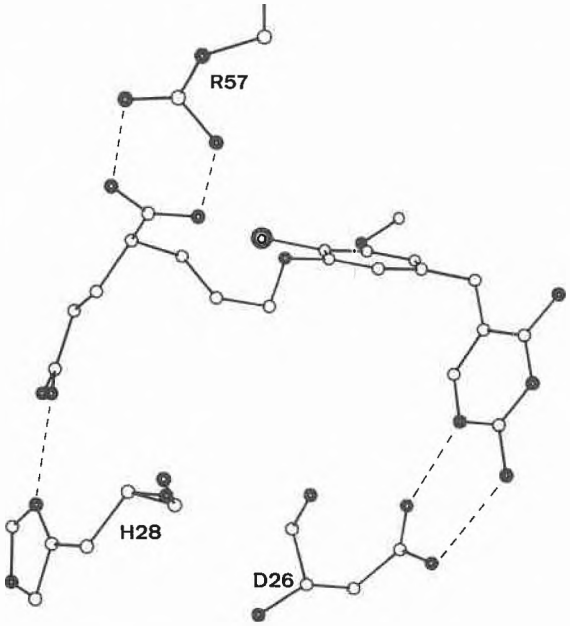


Fig. 4: Conformationally optimized fit of the brodimoprim dicarboxylic acid derivative (Rl = $(CH_2)_3CH(COOH)CH_2CH_2COOH$, Fig. 1) into the *L.casei* DHFR binding site.

The dicarboxyclic acid derivative shown in Fig. 4 is of particular interest. It has been designed [3] as a potentially strongly binding brodimoprim analogue. This derivative can be easily fitted into the *L.casei* DHFR cavity with its two carboxylate groups properly positioned for tight interactions with both Arg57 and His28. Gratifyingly, this compound has been found to bind three orders of magnitude more strongly to the enzyme than brodimoprim and its putative binding mode has recently been corroborated by detailed NMR studies of the binary enzyme-inhibitor complex [4].

In a second application, the natural substrate, dihydrofolate (FAH2), was modeled, using standard strucutral parameters, and introduced into *L.casei* DHFR with its dihydropteridine ring inversely matched onto the pteridine ring of MTX [2]. There is a large body of evidence [7-9] that suggests such an inverse mode of binding for the natural substrate. The interesting question is, whether in this new orientation the p-aminobenzoyl-glutamate moiety can be matched onto the corresponding moiety of MTX in its given conformation, and, in particular, whether the α-carboxylate of glutamate can still reach Arg57. This problem is examined in three steps:

Allowing for all torsions degrees of freedom in FAH2 (except rotation about the amide bond), but keeping the orientation of the dihydropteridine unit fixed, no satisfactory match is obtained (Fig. 5). The benzoyl and carboxylate groups of FAH2 remain displaced by 1.7-2 A and 0.7-0.9 A, respectively, from the corresponding groups in MTX and no satisfactory interaction of the α-carboxylate with the guandinium unit of Arg57 can be obtained.

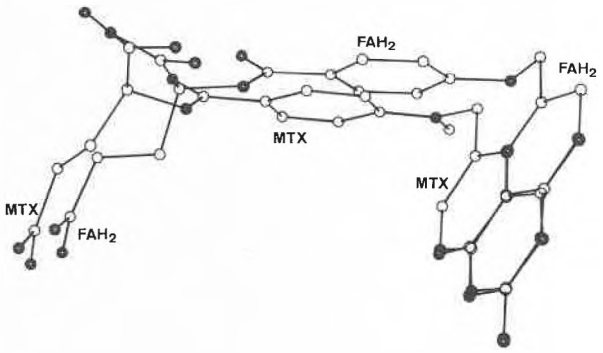


Fig. 5: Superposition of dihydrofolate onto methotrexate with the dihydropteridine ring inversely matched to the pteridine ring.

With the same torsional flexibility, but the additional freedom of overall rotation and translation for FAH2, a much improved fit can be enforced for the glutamate unit. However, in this conformation (Fig. 6, Structure a) the dihydropteridine ring is markedly shifted and rotated out of its optimum orientation for tight contacts to Asp26.

An almost perfect match of both the glutamate side chain and the pteridine units of FAH2 and MTX can be achieved, if, in addition to torsional flexibility and overall rotation and translation, configurational flexibility is assigned to the olefinic dihydropteridine ring carbon atom, C6, carrying the aminobenzoyl-glutamate side chain. The resulting structure (Fig. 6, Structure b) fits into the enzyme cavity with virtually no spatial conflicts, while maintaining tight contacts to both Asp 26 and Arg57. Interestingly, the degree of

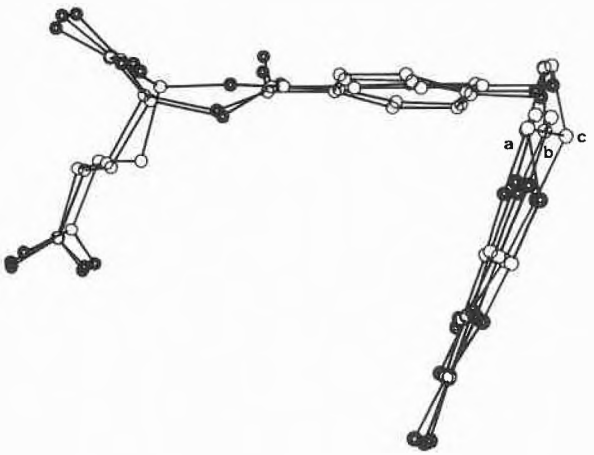


Fig. 6: Superposition of dihydrofolate (structures a and b) and tetrahydrofolate (structure c), conformationally optimized to fit into the active site of L.casei DHFR.
Structure a: FAH2 with fixed trigonal planar configuration at C6; structure b: FAH2 with partial pyramidalization at C6.

pyramidalization at C6 is about halfway between trigonal planar and tetrahedral.
Application of the same strategy to tetrahydrofolate (FAH4), the enzyme reaction product, leads to structure c (Fig. 6). The overall fit into the protein cavity is less favorable due to the more pronounced puckering of the tetrahydro-pyrazin moiety and the higher degree of pyramidalization at C6.
Based on purely geometrical criteria — optimum orientation of the pteridine unit with respect to Asp26, tight symmectrical contact of the α-carboxylate with the guanidinium unit of Arg57, spatial constraints within the cavity of *L.casei* DHFR — structure b, which may resemble the transition state for hydrogen transfer to C6, fits better into the active site than either the undeformed enzyme substrate (structure a) or product (structure c). This conforms to the concept of 'enzyme catalysis by transition state stabilization'. Partial pyramidalization at C6 activates the C6 = N5 double bond. Thus, Arg57, a conserved residue in all known DHFR, may play the role of a remote hydrogen transfer activator by inducing C6 paramidalization through interaction with the α-carboxylate group of the FAH2 glutamate unit.
Clearly, this argument rests upon the assumption of a (partially) rigid protein. We are presently not in a position to judge the amount of flexibility in *L.casei* DHFR. However, raster analyses of residue packing reveal that Arg57 and Asp26 (*L.casei* DHFR) as well as Arg57 and Asp27 (*E.coli* DHFR) are tightly embedded by neighboring residues, with only their terminal heteroatoms exposed at the back of narrow entry channels for the pteridine and carboxylate groups, respectively. A conformational adjustment of the protein to accommodate an undeformed substrate molecule would therefore seem to require cooperative motions of extended peptide domains.
In a third application, the possibility of fitting the cofactor NADPH into the rigid *E.coli* DHFR-MTX binary complex was examined. Despite only limited sequence homology between *L.casei* and *E.coli* DHFR (27% identity), the overall three-dimensional architectures of the two proteins are remarkably similar. Principal axes orientation of the two proteins leads to almost perfect superposition of major parts of their Cα-backbone. Cavity analyses for the binary complex revealed a substantial amount of unused space next to the site occupied by MTX.

To probe a potential fit of NADPH into this binding site, a suitable model of NADPH was first constructed using our model building facilities. Relevant substructural units from the Cambridge Structural Database [11] were extracted, combined, and supplemented using standard bond lengths and valence angles. This model was then introduced into the binary *E.coli* complex so that the spatial relationship between the nicotinamide and pteridine rings was the same as in the ternary *L.casei* complex. In this position, the nicotinamide unit just fills the unused half-space in a deep cavity that is partly occupied by the MTX pteridine ring. The carboxamide group is properly situated for interactions with the protein backbone (Ala7 and Ile14) via three hydrogen bonds. The binding of the nicotinamide unit is thus very similar to the one found in the ternary *L.casei* complex [10].

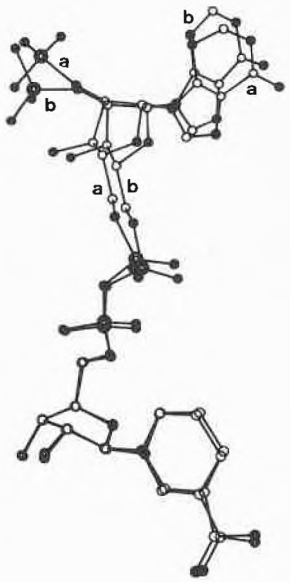


Fig. 7: Superposition of two NADPH molecules.
Structure a: NADPH, as retrieved from the structural data of the ternary *L.casei* DHFR-MTX-NADPH complex [2].
Structure b: NADPH, as modeled into the binary *E.coli* DHFR-MTX complex.

With its nicotinamide unit fixed in space, NADPH was conformationally relaxed with respect to all tor-

sional degrees of freedom in order to minimize spatial conflicts with the (rigid) protein. Only minor changes in torsion angles were necessary in order to produce a fitting conformation, as can be seen by the superposition of the resultant NADPH structure (Fig. 6, structure b) and the experimentally determined NADPH conformation (Fig. 6, structure a) retrieved from the ternary *L.casei* complex [2]. Major conformational changes are predicted to occur mainly in the adenosin-2'-phosphate unit resulting in modified relative orientations of the adenine ring and the 2'-phosphate group.

Comparative raster analyses of the packing of NADPH to DHFR for both ternary complexes reveal striking similarities between the experimental *L.casei* and the hypothetical *E.coli* cases. The most notable differences occur at the enzyme surface where NADPH is 'wrapped' by interacting residues in the *L.casei* case, but is more exposed in the *E.coli* case. Thus, it appears that the cavity for NADPH is essentially preformed in the binary *E.coli* DHFR-MTX complex and that binding of NADPH will not induce major conformational changes in the protein, except for conformational adjustments of mainly peripheral amino acid residues.

Acknowledgement

I wish to thank my coworkers, Dr. *H.-J. Ammann* and *G. Schrepfer*, who have contributed much to the successful development of our new computer graphics system, and my colleague, Dr. *I. Kompis*, for many stimulating discussions.

References

1 *K, Mueller, H.-J. Ammann* and *G. Schrepfer:* unpublished results.
2 *J.T. Bolin, D.J. Filman, D.A. Matthews, R.C. Hamlin* and *J. Kraut:* J. Biol. Chemistry *257,* 13650 (1982).
3 *I. Kompis* and *R. Then:* submitted for publication.
4 *I. Kompis, R.L. Then, K. Mueller, A. Kroehn, B. Birdsall, J. Feeney, C. Pascual* and *G.C.K. Roberts:* J. Med. Chem., in press.
5 *L.F. Kuyper, B. Roth, D.P. Baccanari, R. Ferone, C.R. Beddell, J.N. Champness, D.K. Stammers, J.G. Dann, F.E.A. Norrington, D.J. Baker* and *P.J. Goodford:* J. Med. Chem. *25,* 1120 (1982).
6 *T.F. Koetzle* and *G.J.B. Williams:* J. Amer. chem. Soc. *98,* 2074 (1976).
7 *J.E. Gready:* Adv. Pharmacol. Chemother. *17,* 37 (1980).
8 *P.A. Charlton, D.W. Young, B. Birdsall, J. Feeney* and *G.C.K. Roberts:* J.C.S. Chem. Commun. 922 (1979).
9 *J.C. Fontecilla-Camps, C.E. Bugg, C. Temple, Jr., J.D. Rose, J.A. Montgomery* and *R.K. Kisliuk:* J. Amer. chem. Soc. *101,* 6114 (1979).
10 *D.J. Filman, J.T. Bolin, D.A. Matthews* and *J. Kraut:* J. Biol. Chemistry *257,* 13663 (1982).
11 *F.H. Allen, S. Bellard, M.D. Brice, B.A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B.G. Hummelink-Peters, O. Kennard, W.D.S. Motherwell, J.R. Rodgers* and *D.G. Watson:* Acta Cryst. B35, 2331 (1979).